Распространение радиоволн

Propagation of Radio Waves

Распространение радиоволн

M. Dolukhanov

URSS

Propagation of Radio Waves

М. Долуханов
РАСПРОСТРАНЕНИЕ РАДИОВОЛН
Москва, "Связь"

M. Dolukhanov

Propagation of Radio Waves

URSS
MOSCOW
1995

М. Долуханов
РАСПРОСТРАНЕНИЕ РАДИОВОЛН
Москва, "Связь"

ISBN 5-88417-082-3

© M. Dolukhanov, 1965.
© Translation, Boris Kuznetsov, 1971.
© "УРСС", 1995.

Contents

Chapter One. AN OUTLINE OF RADIO WAVE PROPAGATION . . . 9

1.1. Basic Definitions . . . 9
1.2. Classification of Radio Waves According to the Mechanism of Propagation . . . 13
1.3. Units and Symbols . . . 18
1.4. Free-space Propagation of Direct Waves . . . 21
1.5. Attenuation of Radio Waves Propagated in an Imperfectly Conducting Medium . . . 26
1.6. Reflection of Radio Waves from the Earth's Surface . . . 31
1.7. Attenuation Function. Design Principles for Radio Links with Time-invariant Parameters . . . 37
1.8. Signal Distortion in Propagation. Design Principles for Radio Links with Time-varying Parameters . . . 39

Chapter Two. GROUND-WAVE PROPAGATION . . . 41

2.1. Electrical Properties of the Earth's Surface . . . 41
2.2. Ground-wave Propagation over a Flat Earth with Elevated Transmitting and Receiving Aerials . . . 43
2.3. Leontovich's Approximate Boundary Conditions . . . 60
2.4. The Field Structure at the Point of Reception . . . 63
2.5. Radio Propagation over the Flat Earth, with the Aerials Located Directly at the Interface . . . 70
2.6. Radio Propagation over an Inhomogeneous Surface. Coastal Refraction . . . 80
2.7. Spherical-earth Problem. Line-of-Sight Propagation. Effect of the Earth's Curvature . . . 93
2.8. Radio Propagation over a Smooth Homogeneous Spherical Earth . . . 100
2.9. Radio Propagation over an Inhomogeneous Spherical Earth 117
2.10. Significant Volumes of Radio Transmission . . . 118
2.11. Line-of-Sight Radio Propagation over Hills . . . 124
2.12. Line-of-Sight Radio Propagation over a Rough Surface . . . 129
2.13. Radio Propagation over a Knife Edge . . . 133

Chapter Three. THE EFFECT OF THE TROPOSPHERE ON GROUND WAVES. PROPAGATION OF TROPOSPHERIC WAVES . . . 140

3.1. General Properties of the Troposphere. Composition and Structure . . . 140
3.2. Refractive Index of the Troposphere . . . 142

3.3. The Effect of the Troposphere on Ground Waves. Atmospheric Refraction 144
3.4. Effective Earth's Radius 146
3.5. The Effect of Fluctuations in the Troposphere on Ground Wave Propagation 150
3.6. Forms of Atmospheric Refraction 151
3.7. Conditions for Super-refraction 155
3.8. Radio Propagation under Conditions of Super-refraction . . . 157
3.9. Long-distance Propagation of Ultra-short Waves by Tropospheric Scattering 159
3.10. Determination of the Effective Scattering Cross-section and the Significant Scattering Volume in the Troposphere 167
3.11. Fading in Tropospheric Scatter Propagation 174
3.12. Seasonal Variations in the Signal Level in Tropospheric Propagation of Ultra-short Waves. The Effect of Climatic Conditions 188
3.13. Approximate Calculations of Tropospheric Radio Links 191
3.14. Attenuation of Radio Waves in the Troposphere 195
3.15. Tropospheric Propagation of Radio Waves at Optical Frequencies 206

Chapter Four. THE IONOSPHERE 209

4.1. The Constitution and Structure of the Upper Atmosphere 209
4.2. The Mechanism of Ionization 215
4.3. Ionizing Agencies 218
4.4. Ionization in a Homogeneous Atmosphere 220
4.5. Disappearance of Free Charges in the Atmosphere 225
4.6. Ionization in the Real Atmosphere 228
4.7. Radio Propagation in a Homogeneous Ionized Gas 235
4.8. The Phase and Group Velocities of Radio Waves in an Ionized Gas 238
4.9. Radio Propagation in a Homogeneous Ionized Gas in the Presence of a Permanent Magnetic Field 242
4.10. Refraction and Reflection of Radio Waves in the Ionosphere 250
4.11. Ionospheric Stations 258
4.12. Virtual Ionospheric Height and Path 263
4.13. Ionospheric Maps 267

Chapter Five. IONOSPHERIC PROPAGATION 271

5.1. Physical Processes in the Propagation of Sub-audio Waves. Magnetohydrodynamic Waves 271
5.2. Physical Processes in the Propagation of Audio-frequency Waves. Whistling Atmospherics 274
5.3. Physical Processes in the Propagation of Very Long and Long Waves 275
5.4. Variations in VLW and LW Propagation 279
5.5. Calculation of Field Strength in the LW and VLW Bands 281
5.6. Physical Processes in the Propagation of Medium Waves . . 284
5.7. Variations in MW Propagation 285
5.8. Calculation of the Sky-wave Signal in the MW Band . . . 289
5.9. Physical Processes in the Propagation of Short Waves . . . 290
5.10. Fading in the SW Band 295
5.11. The Primary Skip Zone 299
5.12. Short-wave Echoes 301

5.13. The Sunspot-cycle Variations in Short-wave Signal Strength 304
5.14. The Effects of Geomagnetic Disturbances on Short-wave Propagation 307
5.15. Calculation of Short-wave Radio Circuits 310
5.16. Propagation of Metric Waves Due to Reflections from the Regular Regions of the Ionosphere and the Sporadic *E* Layer 323
5.17. Propagation of Metric Waves Through Scatter from the Ionosphere 324
5.18. Propagation of Metric Waves Due to Meteor Trails 328
5.19. Modes of Propagation for Metric Waves 332
5.20. Propagation of Decimetric and Centimetric Waves 333
5.21. Propagation of Millimetric Waves 333
5.22. Propagation of Optical-frequency Radio Waves 336

Chapter Six. ATMOSPHERIC AND COSMIC NOISE 337

6.1. Main Sources of Noise 337
6.2. Atmospheric Noise 337
6.3. Cosmic Noise 344
6.4. Thermal Noise Due to the Atmosphere and Earth's Surface 349

Chapter Seven. SPACE COMMUNICATION 351

7.1. Basic Definitions 351
7.2. General Principles of Satellite Communications 353
7.3. Energy Considerations in Satellite Communications Systems 356
7.4. Prospects of World-wide Satellite Communications Systems 359
7.5. Ground-to-Space and Space-to-Space Communications . . . 360

References . . 362
Index 366

Chapter One

An Outline of Radio Wave Propagation

1.1. Basic Definitions

Radio waves is a term arbitrarily applied to electromagnetic waves in the frequency range from 0.001 to 10^{16} hertz. In terms of wavelength, the lower limit of radio waves propagated in free space (a vacuum) is 3×10^{11} metres, and the upper limit, 3×10^{-8} metres.

It is relevant to note that these limits of the radio-wave spectrum are considerably wider than they were assumed to be until quite recently. On the low-frequency side, it was customary to set the limit at an audio frequency of 10^3 hertz. Incidentally, this is the limit for the frequencies used by transmitters. Recent research has shown, however, that some natural phonomena involve radio waves with frequencies down to a few thousands of a hertz. Such frequencies, are, for example, generated by the fluctuations of the solar electron-proton stream as it penetrates the earth's atmosphere. These waves are closely related to what have come to be called *magneto-hydrodynamic waves* —mechanical waves produced in the ion plasma of the atmosphere. They are also produced by lightning discharges.

Obviously, frequency 0.001 hertz is an arbitrary limit, for, as science advances, one will have to deal with progressively lower radio frequencies.

On the high-frequency side, the radio-wave spectrum was once limited to 10^{12} hertz, the *sub-millimetre range*. With the advent of lasers, it has become possible to generate waves in the optical range, including the ultra-violet, by radio methods. With every year, the upper boundary of laser frequencies is rolled back more and more, so that frequency 10^{16} hertz is likewise an arbitrary limit which will have to be revised before long.

In the Soviet Union, it is customary to break down the entire radio spectrum into bands on a decimal basis, with the bands named according to the wavelengths they encompass. The longest and shortest waves have not yet been named, and the former will tentatively be referred to as sub-audio and audio radio waves in this text. The subdivision of radio waves into bands is presented in Table 1.1.

Table 1.1. THE SPECTRUM OF RADIO FREQUENCIES

Frequency range (lower limit exclusive, upper limit inclusive)	Wavelength range (lower limit exclusive, upper limit inclusive)	Description	International band designations
1	2	3	4
3 mHz-3 kHz	10^{11}-10^{5} m	Sub-audio and audio radio waves	Extra-low frequencies (ELF)
3-30 kHz	10^{5}-10^{4} m	Very long waves (VLW)	Myriametric waves or very low frequencies (VLF)
30-300 kHz	10^{4}-10^{3} m	Long waves (LW)	Kilometric waves or low frequencies (LF)
300-3000 kHz	10^{3}-10^{2} m	Medium waves (MW)	Hectometric waves or medium frequencies (MF)
3-30 MHz	10^{2}-10 m	Short waves (SW)	Decimetric waves or high frequencies (HF)
		Ultra-short waves (USW):	
30-300 MHz	10-1 m	Metric waves	Metric waves or very high frequencies (VHF)
300-3000 MHz	1 m-10 cm	Decimetric waves	Decimetric waves or ultra-high frequencies (UHF)
3-30 GHz	10-1 cm	Centimetric waves	Centimetric waves or super-high frequencies (SHF)
30-300 GHz	1 cm-1 mm	Millimetric waves	Millimetric waves or extremely-high frequencies (EHF)
300 GHz-300 THz	1 mm to 0.75 micron	Optical radio waves, infra-red	—
400 THz-750 THz	0.75 to 0.4 micron (7500 to 4000 Å)	Visible light	—
750-3000 THz	0.4 to 0.1 micron (4000 to 1000 Å)	Ultra-violet	—

A few remarks should be added to the table. To begin with, waves longer than 10^5 metres, that is, audio and sub-audio radio waves have not yet found commercial uses and are employed solely in research.

In the Soviet literature on the subject, the wavelength bands are usually designated by their abbreviations, СДВ for very long waves, ДВ for long waves, СВ for medium waves, КВ for short waves, МВ for metric waves, ДМВ for decimetric waves, СМВ for centimetric waves, and ММВ for millimetric waves. The infra-red and ultra-violet are designated ИКЛ and УФЛ, respectively. Wavelengths from 0.1 micron to 1 millimetre are referred to as optical radio waves.

Wavelengths in the range from 1 centimetre to 10 metres are often referred to as ultra-short (abbreviated to УКВ in Russian, that is, USW). Sometimes, centimetric and decimetric waves are collectively called *microwaves*, although the same term is often extended to cover the whole of the USW band. It may be added in passing that this is not a particularly happy choice of term.

In view of the advances in the use of optical radio waves, it is convenient to express the respective wavelengths in microns (1 micron = 0.000001 metre = 0.001 millimetre). The still shorter radio waves can be measured in Ångstrom units (1 Ångstrom = 0.000,000,000,1 metre = 0.000,000,1 millimetre = 0.0001 micron).

Radio frequencies are expressed in the following derived units:

1 millihertz (mHz) = 0.001 hertz
1 kilohertz (kHz) = 1000 hertz
1 megahertz (MHz) = 1,000,000 hertz
1 gigahertz (GHz) = 1,000,000,000 hertz
1 terahertz (THz) = 1,000,000,000,000 hertz

It is important to note the difference between the abbreviations for the millihertz (mHz) and the megahertz (MHz).

This text will be concerned with the unguided propagation of radio waves, that is, in the terrestrial atmosphere, over and through the earth, and in outer space. In other words, we shall leave out the propagation of radio waves directed by any man-made structure, such as wire systems, coaxial lines, waveguides, strip or single-wire lines.

Unguided propagation of radio waves (in the sense defined above) finds many and extremely important uses in science and engineering. Above all, mention should be made about the transmission of intelligence over short and long distances (telegraphy, telephony, facsimile, and television); the detection and ranging of objects (radar); control of machinery from a distance (telecontrol); navigational use of radiolocation for determining position or direction of craft (radio navigation); and distance measurements by radio means. Unguided radio propagation is also used in geophysics, in

the study of the upper atmosphere, radio astronomy, the study of the Sun, stars and nebulae inside and outside our Galaxy.

Despite the differences, all of the above applications of unguided radio propagation have much in common. The most important common aspect is that all use a *radio circuit* or *link*, consisting of a transmitter, a receiver, and a propagation medium. In line transmission, the propagation medium is constituted by wires or cables.

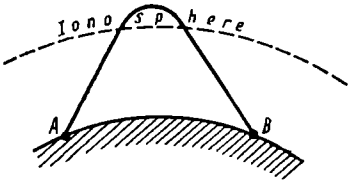


Fig. 1.1. A simple radio link or circuit:
A — transmitter; *B* — receiver. The full line shows the path of radio propagation

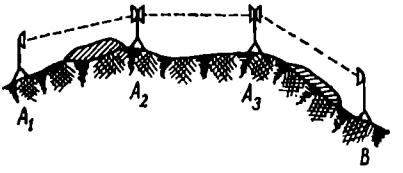


Fig. 1.2. A microwave radio link:
A_1, *B* — terminal stations; A_2, A_3 — repeaters

In radio transmission, it is the natural medium such as the atmosphere around the earth or outer space.

Radio engineering uses three types of radio circuit or link. The simplest one is shown in Fig. 1.1, with the transmitter and receiver located at its terminal points. In the example shown, transmitted waves reach the point of reception due to reflection from the ionosphere. In other cases radio waves can reach the receiver through diffraction round the globe, scattering in the troposphere, or in some other way. Another example is the radio circuit or link between an earth-based station and a space probe. A similar radio circuit or link will be produced by a radio-emitting star as transmitter and a radio telescope as receiver.

Fig. 1.2 shows a *line-of-sight radio relay system*. In this radio link, the message sent out by the transmitter reaches the final receiver after it has passed through a chain of intermediate stations. This is done in cases where messages cannot be transmitted directly to the terminal point because of, say, large distances (AB) involved. Then the entire system is broken down into line-of-sight sections, or *hops* (A_1A_2, A_2A_3, etc.) each of which may be treated as a simple radio circuit. All intermediate stations are *active*, that is, they are receive-amplify-retransmit signal repeaters. At each repeater, the r.f. signal is converted back to the i.f. band, amplified, translated to a microwave frequency, and retransmitted in the direction of the next repeater at a frequency different from the received one.

The third type of radio circuit or link is shown in Fig. 1.3. It is used for *behind-*(or *over-*)*the horizon transmission* and is included in the class of *scatter systems.* In them, the transmitted signal does not reach the receiver B directly. The signal generated by the transmitter at A illuminates a man-made or a natural body, C, which differs from the surrounding medium in electrical characteristics. As a result, the body at C reflects or scatters the incident radiation, and the reflected or forward-scattered signal is picked up at B. In Fig. 1.3. the reflected or scattered radio waves are shown by the dotted lines.

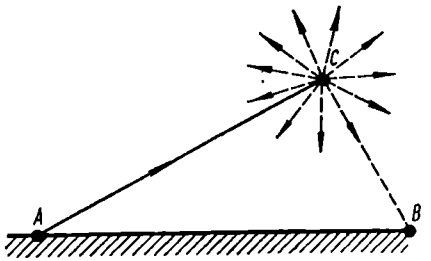


Fig. 1.3. A scatter radio link:
A — transmitter; B — receiver; C — scattering volume

Scatter or behind-the-horizon systems may use tropospheric scatter, ionospheric scatter, meteor trail scatter, reflection from the Moon or artificial earth satellites used as passive reflectors. In the first four cases, scatter is produced by a natural medium, while in the last this is done by a man made object.

In some cases, a scatter system may have its transmitter and receiver located together. Examples are a radar equipment (when body C is the target or targets), and a vertical ionospheric sounding equipment (when body C is the reflecting area in the ionosphere).

1.4. Classification of Radio Waves According to the Mechanism of Propagation

In the past, when radio stations were located either on the ground or a short distance above it, modes of wave propagation were classed according to how the waves travel along the earth's surface. Now that much headway has been made in the practical use of satellite communication, this classification must be extended to include radio propagation through outer space.

The simplest case of radio propagation is the transmission of a wave between a transmitter and a receiver in free space. By "free space" is meant a region whose properties are isotropic, homogeneous, and loss-free. In a first approximation, this definition applies to outer space. In free space, radio waves are propagated in straight lines, and so we may assume that this mode of propagation also holds for outer space. Let us agree to call this a *direct wave.* The path of a direct wave is shown in Fig. 1.4 where A is a transmitter on the ground, and M is a space object. Obviously, communication between

two space objects will also use direct waves. Radio astronomy, too, uses direct waves coming from a star or a nebula to the aerial of a radio telescope. In some cases, though, the direct radio beam can be bent, or diffracted, as it passes through the atmosphere.

Definition: A direct wave is that which propagates through an isotropic and homogeneous medium, notably outer space, in straight lines. In space communication, the direct wave may experience

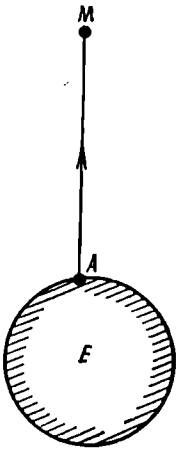


Fig. 1.4. Propagation path of a direct wave:
E — earth; *A* — transmitter; *M* — space object

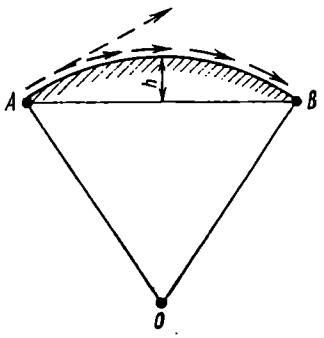


Fig. 1.5. Diffraction of radio waves around the earth's surface

slight diffraction, scattering, and rotation of the plane of polarization as it passes through the earth's atmosphere.

In most cases, however, the terminal points of a radio circuit are located on or near the earth's surface. Under the circumstances, we cannot ignore the effects of the earth and its atmosphere on radio propagation. Detailed studies have shown that radio propagation along the earth's surface is affected by at least three factors. These factors are (1) the proximity of ground and the spherical shape of the earth; (2) inhomogeneity of the troposphere (the lower layer of the earth's atmosphere, extending out to about 15 kilometres); and (3) effects of the ionosphere (the part of the earth's outer atmosphere extending from 60 to about 600 kilometres, where sufficient ions and electrons are present to affect the propagation of radio waves).

Let us consider each of these factors separately.

To show the effect of the earth's proximity in its pure form, we shall assume that the globe is in vacuum. On the one hand, the par-

tially conducting surface of the earth attenuates radio waves (as they penetrate it) and distorts the waveform (as compared with free space propagation). On the other hand, the spherical form of the earth causes the radio waves propagated along its surface to diffract.

From physics it is known that diffraction is noticeable when the obstacle in the way of waves is comparable in size with the wavelength. In our case, the obstacle is the height h of the spherical segment (Fig. 1.5) cut off by the plane normal to the plane of the drawing and passing through the chord AB connecting the point of transmission A and the point of reception B. Table 1.2 gives the values of h for different distance between transmitter and receiver.

Table 1.2. HEIGHT OF THE SPHERICAL SEGMENT FOR DIFFERENT PROPAGATION PATH LENGTHS

Length of propagation path, km	1	5	10	50	100	500	1000	5000
Height of spherical segment, m	0.031	0.78	3.1	78	310	7800	3.1×10^4	3.75×10^5

The value of h is so much greater than the length of optical-frequency radio waves that these do not practically diffract on the earth and are propagated only along the direct ray path. In contrast, metric and kilometric radio waves are diffracted. Their path is diagrammatically shown in Fig. 1.5 by the full arrows, while the broken arrows show the path followed by non-diffracted waves; as is seen, it is along a straight line tangent to the earth's surface at point A.

Diffraction of a radio wave occurs when the wave is propagated across an obstacle. The earth may be regarded as a spherical obstacle, and radio waves are diffracted around the curvature of the globe. However, even under the most favourable conditions, the radio propagation due to diffraction covers distances not exceeding 3000 or 4000 kilometres.

Definition: A radio wave propagated close to the earth's surface and partly following the curvature of the globe due to diffraction is called a *ground* or *surface wave*. In this text, the former term will be used.

Let us now drop the assumption about the earth being placed in a vacuum and suppose that the globe is surrounded only by the troposphere extending out to an altitude of about 15 kilometres, and that past the troposphere there is free space.

As will be shown shortly, the troposphere is an inhomogeneous medium with time-varying properties due to the weather conditions. Its refractive index gradually decreases with height and also shows local irregularities.

Gradual variations in the refractive index result in a bending of the paths taken by radio waves, so that they can follow the curvature of the earth. Thus, the troposphere affects the propagation of ground waves and can promote communication over large ranges.

Local inhomogeneities in index of refraction continually change with time and space, that is, continually vary in configuration and location—as some inhomogeneities disappear, others take their place. Local inhomogeneities cause the scattering of radio waves, and scattered waves can cover ranges up to 1000 kilometres away from the transmitter (Fig. 1.6). This scattering manifests itself only on wavelengths shorter than 10 metres which are only slightly diffracted around the earth's surface. On longer waves, diffraction is more noticeable. Besides, radio propagation due to ionospheric scattering comes into play, and scattered waves become negligible in comparison. Sometimes, weather conditions can produce areas having a steeper decrease of refractive index with height in the lower troposphere. As a result, the atmosphere forms a *duct*, or *waveguide*, that guides the waves shorter than 3 metres (which are diffracted still less) over the surface on ranges up to 800-1000 kilometres from the transmitter.

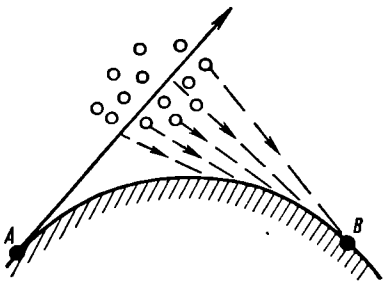


Fig. 1.6. Scattering of radio waves from local irregularities in the troposphere:

A — transmitter; *B* — receiver

Definition: A radio wave propagated over large ranges (up to 1000 kilometres) by tropospheric scattering and ducting is called a *tropospheric wave*. The tropospheric mode of propagation applies only to waves shorter than 10 metres.

Let us now drop the last limitation and assume that the earth is surrounded by its actual atmosphere, including the outer envelope beyond the troposphere.

It has already been noted that the outer part of the atmosphere extending out from 60 to 600 kilometres is called the ionosphere. With respect to radio waves, it acts as an imperfectly conducting surface which reflects radio waves. This reflection from the ionosphere, however, happens only to waves longer than 10 metres. In other words, the ionosphere is opaque to waves longer than about 10 metres,

and they cannot ordinarily leave the atmosphere.* To the shorter waves, including those in the radio-optical band, the ionosphere is a transparent medium.

After a single reflection from the ionosphere, radio waves can cover distances not exceeding 4000 kilometres. However, the waves reflected from the ionosphere can bounce off the imperfectly conducting surface of the earth to be again reflected from the ionosphere. Owing to those multiple reflections from the ionosphere and earth, radio waves can reach localities at any distance from the transmitter and even girdle the earth more than once. Fig. 1.7 shows the propagation of radio waves due to triple reflection from the ionosphere.

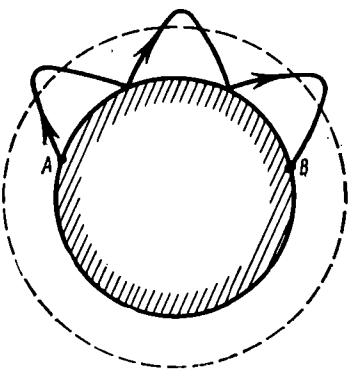


Fig. 1.7. Radio propagation through multiple reflections from the ionosphere

Apart from reflection, the ionosphere can cause the scattering of radio waves owing to minute inhomogeneities in refractive index. Furthermore, meteors impinging on the terrestrial atmosphere produce meteor trails which can also cause the scattering of radio waves. It should be noted that ionospheric and meteor-trail scattering occurs only to waves shorter than 10 metres, that is, the waves which cannot be reflected from the ionosphere in a regular fashion. Ionospheric and meteor-trail scattering can cause metric waves to cover distances up to 2000 kilometres. On wavelengths longer than 10 metres, the effect of ionospheric and meteor-trail scattering is masked by the regular reflection from the ionosphere which produces stronger fields at the point of reception.

Like the troposphere, the ionosphere is a medium of continually varying properties. Superimposed on the slower variations are the faster random variations, or fluctuations. Among other things, these fluctuations affect local inhomogeneities in the ionosphere.

Definition: Radio waves propagated over large ranges and capable of single or multiple reflections from the ionosphere (waves longer than 10 metres), and also waves scattered from inhomogeneities in the ionosphere and reflected from the ionized trails of meteors (metric waves) are called *ionospheric* or *sky waves*. In this text, we shall use the former term.

---

* The only exception is extra-long waves propagated along the lines of force of the earth's magnetic field.

Students without a previous knowledge of radio propagation usually face some difficulty in telling which wavelengths are propagated by the ground, tropospheric or ionospheric mode. The general remarks that follow will help them out.

On short ranges from a transmitter, waves in all bands will be propagated as ground waves.

Since the troposphere affects waves shorter than 10 metres, only metric, decimetric and centimetric waves will be propagated by the tropospheric mode.

It has been noted that regular reflection from the ionosphere occurs only to waves longer than 10 metres. Therefore, ionospheric propagation holds only for very long (VLF), long (LF), medium (MF) and short (HF) waves. Metric (VHF) waves, too, can be propagated as ionospheric ones due to ionospheric and meteor-trail scattering.

Finally, all waves shorter than 10 metres, including millimetric (EHF) and optical waves, can be propagated only along the direct ray path (both in the terrestrial atmosphere and in outer space).

The above remarks are summarized in Table 1.3.

Table 1.3. MODES OF RADIO PROPAGATION

Wave bands	Short range*	Large range	
		day	night
Sub-audio and audio (ELF)	Ground and ionospheric	Ionospheric	Ionospheric
very long (VLF)	Same	Same	Same
long (LF)	Same	Same	Same
medium (MF)	Ground	Ground	Same
short (HF)	Same	Ionospheric	Same
metric (VHF)	Same	Direct, tropospheric, ionospheric	Direct, tropospheric, ionospheric
decimetric (UHF)	Same	Direct, tropospheric	Direct, tropospheric
centimetric (SHF)	Same	Same	Same
millimetric (EHF)	Direct	Direct	Direct
Optical band	Same	Same	Same

* A short range is about 200 kilometres along the earth's surface, except on short and ultra-short waves where the figure is up to 100 kilometres.

1.3. Units and Symbols

This text uses SI (Système Internationale) units which for electromagnetic quantities are the same as the rationalized MKSA units. In all cases, the time dependence is of the form $e^{i\omega t}$.

We shall use two groups of symbols: (a) for the quantities defining the field and source of radio waves; and (b) for the quantities defining the propagation medium.

The field of a radio wave is defined in the following terms:

E = instantaneous electric field strength (or simply field) of a radio wave, volts per metre (V/m);

E_m = peak electric field strength (or field) of a radio wave, volts per metre (V/m);

E_{rms} = rms electric field strength (or field) of a radio wave, volts per metre (V/m);

H = instantaneous magnetic field strength (or field) of a radio wave, amperes per metre (A/m);

H_m = peak magnetic field strength (or field) of a radio wave, amperes per metre (A/m);

H_{rms} = rms magnetic field strength (or field) of a radio wave, amperes per metre (A/m);

S = Poynting vector, watts per square metre (W/m^2);

P_1 = power radiated by the transmitting aerial, watts (W);

P_2 = power received by the receiving aerial, watts (W);

I_m = peak aerial current, amperes (A);

I_{rms} = rms aerial current, amperes (A).

The propagation medium is defined as follows:

$\varepsilon_0 = \frac{1}{36\pi} \times 10^{-9}$ = absolute permittivity of free space (a vacuum), farads per metre (F/m)

ε = same of the medium, farads per metre (F/m)

ε' $\varepsilon/\varepsilon_0$ = relative permittivity, or dielectric constant. Since in the electrostatic system $\varepsilon_0 = 1$, so $\varepsilon' = \varepsilon$ in such units

$\mu_0 = 4\pi \times 10^{-7}$ = permeability of free space (a vacuum), henries per metre (H/m)

μ = permeability of the medium, henries per metre (H/m)

μ' μ/μ_0 = relative permeability, having the same absolute value as the permeability in the electromagnetic system

σ = conductivity of the medium, siemens per metre (S/m)

j = current density, amperes per square metre (A/m^2).

For convenience in using design equations and in view of the fact that in addition to the basic MKSA units we shall use their multiples or sub-multiples, the following rules will apply to our further discussion:

(1) all equations will be followed by the units used;

(2) if all quantities entering a given equation are in basic units (metres, seconds, kilograms, amperes, volts, coulombs, henries, farads, etc.), the units used will be omitted;

(3) if at least one of the quantities entering a given equation is in arbitrary units, all quantities in that equation will be supplied with subscripts to indicate the units used.

If applied consistently, these rules will save the reader likely errors and misunderstandings.

Design equations can be greatly simplified by use of complex permittivity.

Maxwell's first equation for imperfectly conducting media in the adopted units may be written as

$$\varepsilon \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E} = \text{curl}\, \mathbf{H} \text{ A/m}^2 \tag{1.1}$$

where the letters in Clarendon type designate vectors.

If the electric field due to a radio wave varies with time harmonically

$$\mathbf{E} = \mathbf{E}_m e^{i\omega t} \text{ V/m} \tag{1.2}$$

E can be readily expressed in terms of its time derivative

$$\mathbf{E} = -\frac{i}{\omega} \frac{\partial \mathbf{E}}{\partial t} \text{ V/m}$$

Substituting the above expression in (1.1) gives

$$\left(\varepsilon - i\frac{\sigma}{\omega}\right) \frac{\partial \mathbf{E}}{\partial t} = \text{curl}\, \mathbf{H} \text{ A/m}^2 \tag{1.3}$$

On comparing (1.3) with Maxwell's equation for an ideal dielectric which can be derived from (1.1) by putting $\sigma = 0$ such that

$$\varepsilon \frac{\partial \mathbf{E}}{\partial t} = \text{curl}\, \mathbf{H} \tag{1.1a}$$

we note that the complex quantity

$$\varepsilon_c = \varepsilon - i\frac{\sigma}{\omega} \text{ F/m}$$

is in effect the permittivity of an imperfectly conducting medium. This quantity is called the *complex permittivity*. In calculations, it is more convenient to use the *relative* (that is, dimensionless) *complex permittivity*, or the *complex dielectric constant* which is obtained by dividing the complex permittivity into ε_0

$$\varepsilon_c' = \varepsilon_c/\varepsilon_0 = \frac{\varepsilon}{\varepsilon_0} - i\,\frac{\sigma}{\omega\varepsilon_0} = \varepsilon' - i\frac{\sigma\lambda 36\pi \times 10^9}{2\pi c} = \varepsilon' - i60\lambda\sigma \tag{1.4a}$$

where $c = 3 \times 10^8$ metres per second is the velocity of light in free space (a vacuum).

Since the units used in (1.4a) are not identified, according to the rules stated above λ should be in metres, and σ in siemens per metre.

As will be recalled, the first term on the left-hand side of (1.1) represents the displacement current density

$$j_d = \varepsilon \frac{\partial E}{\partial t} \text{ A/m}^2 \tag{1.5}$$

while the second term represents the convection current density

$$j_c = \sigma E \text{ A/m}^2 \tag{1.6}$$

When considering the effect of the earth's imperfectly conducting surface on the propagation of ground waves, the above equations can be simplified, if either the displacement or the convection current prevails in the propagation medium. Therefore, it is of practical importance to know the relative magnitudes of both currents. Dividing (1.5) into (1.6), we get

$$\left|\frac{j_d}{j_c}\right| = \left|\frac{j\varepsilon\omega E}{\sigma E}\right| = \frac{2\pi c\varepsilon' e_0}{\lambda\sigma} = \frac{\varepsilon'}{60\lambda\sigma} \tag{1.7}$$

It will be readily seen that the ratio j_d/j_c is defined in terms of the same quantities that enter Eq. (1.4a) for the complex permittivity.

1.4. Free-space Propagation of Direct Waves

Suppose that a transmitter is radiating an output power P_1 watts from an aerial (the one which radiates equally in all directions). Such an aerial is called an isotropic radiator. We further assume that this isotropic radiator is placed in free space, a homogeneous, non-absorbing medium of dielectric constant unity.

Suppose also that a receiver is situated at a distance r metres from the transmitter (Fig. 1.8). As the transmitter radiates equally throughout the sphere surrounding the aerial, the power flux density, or the Poynting vector, at the distance r is

$$S = \frac{P_1}{4\pi r^2} \text{ W/m}^2 \tag{1.8}$$

In the adopted units, the average value of the Poynting vector over a period is

$$S = E_{rms} H_{rms} \text{ W/m}^2 \tag{1.9}$$

the ratio of the electric to magnetic field strength being

$$H_{rms} = \frac{E_{rms}}{120\pi} \text{ A/m} \tag{1.10}$$

where 120π (or 377 ohms) is the wave impedance of free space. Substituting (1.10) in (1.9) gives

$$S = \frac{E_{rms}^2}{120\pi} \text{ W/m}^2 \tag{1.11}$$

Equating (1.8) and (1.11) and solving for E_{rms} yields

$$E_{rms} = \frac{\sqrt{30P_1}}{r} \text{ V/m} \qquad (1.12)$$

Of course, isotropic radiators are non-existent, but they are approached very closely by *directional aerials*, that is, ones concentrating the radiated energy in a desired direction. The advantages accruing from the directional properties of an aerial, or its directivity, are described by the *power gain*, G, over an isotropic radiator.

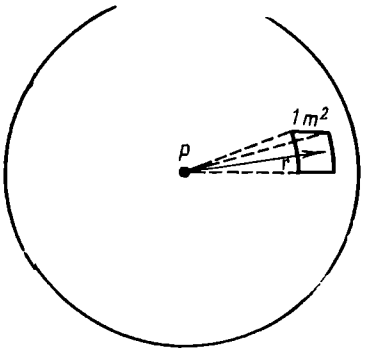


Fig. 1.8. Wave field due to an isotropic radiator

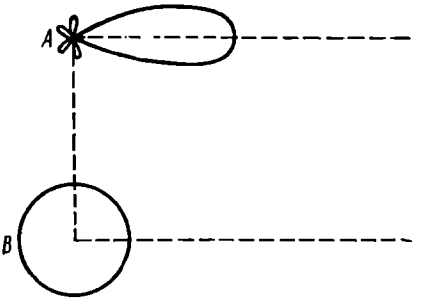


Fig. 1.9. Polar diagrams of (A) a directional aerial and (B) an isotropic aerial

To get an idea about the power gain, let there be a directional aerial A and an isotropic radiator B, placed side by side. Their directional characteristics are presented by *polar diagrams* in Fig. 1.9. As should be expected, the isotropic radiator has a circular polar diagram. If both radiate the same power P_1, it is obvious that the directional aerial will produce a stronger field at a distant receiver. Now, let the power applied to the isotropic source be gradually increased until it produces the same field strength at the receiver as the directional aerial. The factor G_1 showing how many times the power applied to the isotropic radiator should be increased so as to produce the same field strength, is the *power gain* (or *directive gain**) of the aerial.

Thus, in terms of the received field strength a directional aerial is equivalent to an isotropic radiator radiating G_1 times power. Therefore, the equation for the free-space field due to a directional

* The directive gain differs somewhat from the power gain because it allows for the aerial efficiency. We may well ignore this nicety, however, because the efficiency of many directional aerials is close on 100 per cent.

aerial may be written

$$E_{rms} = \frac{\sqrt{30 P_1 G_1}}{r} \text{ V/m} \tag{1.13}$$

where G_1 is the power gain of the directional aerial.

The peak field is then given by

$$E_m = \frac{\sqrt{60 P_1 G_1}}{r} \text{ V/m} \tag{1.13a}$$

and the instantaneous field

$$E = \frac{\sqrt{60 P_1 G_1}}{r} \cos \omega (t - r/c) = \frac{\sqrt{60 P_1 G_1}}{r} \cos (\omega t - kr) \text{ V/m} \tag{1.14}$$

where $k = \omega/c = 2\pi/\lambda$ is the wave number for air.

As often as not, instead of trigonometric notation, Eq. (1.14) is written in symbolic notation

$$E = \frac{\sqrt{60 P_1 G_1}}{r} e^{i(\omega t - kr)} \text{ V/m} \tag{1.14a}$$

Symbolic notation materially simplifies all intermediate derivations, but it is important to remember that the field in question is the real part of (1.14a). For simplicity the symbol Re (for "real part") is often omitted.

The units used in Eqs. (1.12) through (1.14) are inconvenient for practical use — the field is in volts per metre, and the distance in metres. A more convenient form is to express the power in kilowatts, the distance in kilometres, and the field in millivolts per metre. Then

$$E_{rms} = \frac{173 \sqrt{P_{1\text{kW}} G_1}}{r_{\text{km}}} \text{ mV/m} \tag{1.15}$$

and

$$E_m = \frac{245 \sqrt{P_{1\text{kW}} G_1}}{r_{\text{km}}} \tag{1.16}$$

The reader will have noted that the units in Eqs. (1.15) and (1.16) are written as subscripts according to the rules formulated earlier.

At one time, it was customary to evaluate the propagation of radio waves in terms of the electric field produced by the transmitter at the point of reception. This criterion was more or less justified when radio communication mainly used long, medium, and partly short waves. With the growing use of ever shorter wavelengths, it is more reasonable to describe the conditions of propagation in terms of the power produced at the input to the receiver, the more so that receiver sensitivity is now usually stated in terms of the power required for reliable reception. This calls for knowledge of the power gain, G_2, of the receiving aerial. However, the power gain of the transmitting

aerial, G_1, must also be known. Therefore, to avoid consideration of any specific type the power produced in the receiving aerial may be found from the assumption that both aerials are isotropic, that is, $G_1 = G_2 = 1$.

In texts on aerials it is shown that the effective aperture (or area) of an aerial, A, that is, the area over which the aerial gathers the energy of the incoming radio wave, is related to the power gain G as

$$A = \frac{\lambda^2}{\pi} \text{ m}^2 \tag{1.17}$$

To determine the power applied to the receiver input, it will suffice to multiply the power flux density (the Poynting vector) at the point of reception by the effective aperture of the aerial.

For a directional aerial, the Poynting vector is G_1 times the value given by (1.8), that is

$$S_2 = \frac{P_1 G_1}{4\pi r} \text{ W/m}^2 \tag{1.18}$$

or, in more convenient units,

$$S_2 = 8 \times 10^4 \frac{P_{\text{kW}} G_1}{r^2_{\text{km}}} \text{ nW/m}^2 \tag{1.18a}$$

Multiplying by A_2 gives

$$P_2 = S_2 A_2 = \frac{P_1 G_1 G_2 \lambda^2}{(4\pi r)^2} \text{ W} \tag{1.19}$$

or, in more convenient units,

$$P_2 = 6.33 \times 10^3 P_{\text{kW}} G_1 G_2 \frac{\lambda^2_m}{r^2_{\text{km}}} \text{ nW} \tag{1.19a}$$

In the design of radio links, it is convenient to use the concept of *transmission loss*, defined as the ratio of the power radiated from the transmitting aerial to the power available at the terminals of the receiving aerial when both are placed in free space. Then, using Eq. (1.19), we have

$$L_{fs} = P_1/P_2 = \left(\frac{4\pi r}{\lambda}\right)^2 \frac{1}{G_1 G_2} \tag{1.20}$$

where the subscript *fs* identifies free-space propagation.

To give a factor independent of the aerials used, a *basic transmission loss*, L_b, or *path loss*, may be defined. This is the transmission loss between ideal isotropic sources placed in the same position as the actual aerials in free space. Thus, putting $G_1 = G_2 = 1$ in Eq. (1.20), we get

$$L_{b(fs)} = \left(\frac{4\pi r}{\lambda}\right)^2 \tag{1.20a}$$

Since the transmission loss may vary within extremely broad limits, it is usually expressed in decibels, and not in absolute units. Then (1.20) and (1.20a) will take the form

$$L_{fs,\,\mathrm{db}} = 10 \log_{10} L_{fs} = 20 \log_{10} (4\pi r) - 20 \log_{10} \lambda - G_{1\mathrm{db}} - G_{2\mathrm{db}} \quad (1.21)$$

and

$$L_{b,\,\mathrm{db}} = 10 \log_{10} L_{b(fs)} = 20 \log_{10} (4\pi r) - 20 \log_{10} \lambda \quad (1.21a)$$

where the distance and wavelength are in metres.

Equations (1.21) and (1.21a) are convenient also because in aerial design practice the aerial gain is usually expressed in decibels, too.

So far, we have described transmitting aerials in terms of radiated power and gain. On long and medium waves, it is sometimes convenient to do this in terms of aerial current and *equivalent length.*

Let there be an elemental radiator (Fig. 1.10) of length Δl, carrying a current I_{rms}, which is placed in free space so that its axis makes an angle ϑ with the direction of the receiver r distant from the source. Then the electric field at the receiver will be

$$E_{rms} = \frac{60\pi I_{rms} \Delta l}{\lambda r} \sin \vartheta \text{ V/m} \quad (1.22)$$

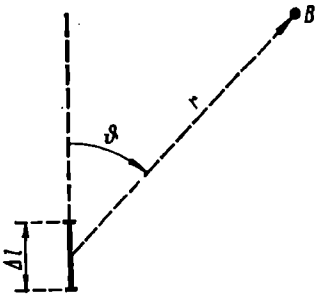


Fig. 1.10. The field due to an elementary radiator

where $\sin \vartheta$ describes the directional properties of the elemental radiator. E_{rms} is a maximum in the equatorial plane when $\vartheta = 90°$. To apply Eq. (1.22) to a finite-length aerial, we replace Δl with the equivalent length, l_e which is the equivalent length as regards radiated power. Quite appropriately, it is also known as the *radiation length.* Then, the field in the equatorial plane is

$$E_{rms} = \frac{60\pi I_{rms} l_{rms}}{\lambda r} \text{ V/m} \quad (1.23)$$

or, in more convenient units,

$$E_{rms} = \frac{60\pi I_{rms_A} l_{em}}{\lambda_{\mathrm{m}} r_{\mathrm{km}}} \text{ mV/m} \quad (1.23a)$$

where I_{rms} is the rms current at an antinode of the aerial.

A radiator of any directivity, when placed in a homogeneous and isotropic medium, produces spherical waves, that is, waves the equiphase surfaces of which form spheres concentric on the radiator. This is obvious since the phase φ of a wave as found from (1.14)

$$\cos \omega (t - r/c) = \cos \left(\omega t - \frac{\omega r}{c} \right) = \cos (\omega t - \varphi)$$

is a function of the distance covere and the velocity of propagation. In an isotropic and homogeneous mlium, the velocity of propagation is the same in all directions, and t condition $\varphi = \omega r/c = \text{constant}$ is the same as $r = \text{constant}$, which the equation of a sphere in spherical coordinates.

Example 1.1. Find the basic tnsmission loss in free space for two cases: (a) $r = 10$ km, $\lambda = 2000$ m and (b) $r = 10^7$ km (space communication), $\lambda = 3$ cm.

Solution: Substituting the numical values of r and λ in (1.20a) and (1.21a) gives:

(a) $L = 39.3$ or 15.9 db

(b) $L = 1.75 \times 10^{25}$ or 252 db

As is seen, the transmission los in practical cases can vary over a broad range. It is also seen thadecibels offer a convenient form of expression in such cases.

1.5. Attenuation of Radio Waves Propated in an Imperfectly Conducting Meam

When radio waves are propagad in free space (including outer space), the decrease in their pwer flux density with distance is caused solely by the effect of spading or divergence, that is, as a wave advances its energy is dispsed on the surface of an expanding sphere.

In most practical cases, radio wves are propagated through media which are imperfect conductors. Tse are sea water, the earth's crust, and the ionized layers of the atmphere. Propagated in such imperfect conductors, radio waves inace currents there, which cause absorption of energy. This entas irrevocable losses of energy as heat, and the radio waves are atnuated.

The extent to which a radio wae is attenuated can be conveniently determined by use of compk permittivity.

From field theory it is known tlt the electric field of a radio wave propagated in an ideal dielectric is described as

$$E_z = E_m e^{i\omega\left(t - \frac{x\sqrt{\varepsilon'}}{c}\right)} \text{ V/m} \qquad (1.24)$$

where ε' is the dielectric constanof an ideal dielectric and $\sqrt{\varepsilon'}/c = 1/u$, or the reciprocal of the vocity of propagation. It is assumed that the wave is propagated in te direction of axis X and the electric field is oriented in the direion of axis Z in the right-handed system of coordinates shown in ig. 1.11.

Formally, Eq. (1.24) may be extended to include an imperfect conductor by replacing the real ε' with the complex $\varepsilon'_c = \varepsilon' - i\,60\,\lambda\sigma$.

Putting

$$\sqrt{\varepsilon'_c} = n - ip \tag{1.25}$$

and substituting in (1.24) gives

$$E_z = E_m e^{-\frac{\omega}{c} px} e^{i\omega\left(t-\frac{xn}{c}\right)} \text{ V/m} \tag{1.26}$$

Eq. (1.26) shows that in contrast to a perfect dielectric which does not absorb the energy of radio waves, radio propagation in an imper-

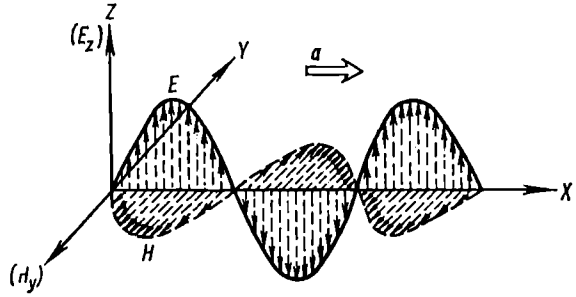


Fig. 1.11. The electric and magnetic fields of a plane radio wave in an ideal dielectric

fect conductor is accompanied by a reduction in the wave amplitude in exponential fashion.

The quantity

$$\delta = \frac{\omega}{c} p \text{ m}^{-1} \tag{1.27}$$

is called the *absorption coefficient*.

From (1.26) it follows that in an imperfect conductor the velocity of a radio wave is

$$u = \frac{c}{n} \text{ m/s} \tag{1.28}$$

that is, the velocity u is a function of the permittivity ε', the conductivity σ, and the frequency. The wavelength, too, is changed in an imperfect conductor. By definition,

$$\lambda' = u/f = c/nf = \lambda/n \text{ m} \tag{1.29}$$

where λ is the wavelength in air.

To determine the numerical values of n and p in the expressions for the absorption coefficient and wavelength in an imperfect con-

ductor, let us square both sides of (1.25)

$$\varepsilon' - i60\lambda\sigma = n^2 - p^2 - i2np$$

This complex equality is equivalent to

$$\left.\begin{aligned} n^2 - p^2 &= \varepsilon' \\ np &= 30\lambda\sigma \end{aligned}\right\} \tag{1.30}$$

Solving (1.30) for n and p and discarding superfluous signs, we get

$$n = \pm \sqrt{\frac{1}{2}\left[\varepsilon' + \sqrt{(\varepsilon')^2 + (60\lambda\sigma)^2}\right]} \tag{1.31}$$

$$p = \pm \sqrt{\frac{1}{2}\left[-\varepsilon' + \sqrt{(\varepsilon')^2 + (60\lambda\sigma)^2}\right]} \tag{1.32}$$

The two signs before the outer radicals indicate that the wave can move to and from a point of reference, as shown in Fig. 1.12 for the

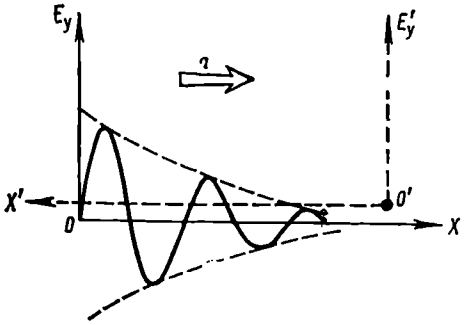


Fig. 1.12. Explaining the duality of signs of n and p

electric field in a strongly absorbing medium. For an observer at point O, the wave will move away, and the amplitude will decay exponentially, $e^{-\delta x}$, with distance from the origin of coordinates; therefore, in Eqs. (1.31) and (1.32) the upper signs should be chosen. For an observer at point O', the wave will approach, and its amplitude will rise exponentially, $e^{\delta x}$, with distance x, therefore, the lower signs should be chosen for n and p.

There is a way of materially simplifying Eqs. (1.31) and (1.32) in practical cases. When the displacement current is many times the convection current, then by (1.7)

$$\varepsilon' \gg 60\lambda\sigma$$

Neglecting the second term under the inner radical of (1.31) on the strength of the above inequality, we get

$$n \approx \sqrt{\varepsilon'} \tag{1.33}$$

This crude approximation is inapplicable to p. Applying the binomial theorem to the inner radical of (1.32) gives

$$p = \sqrt{-\frac{\varepsilon'}{2} + \frac{\varepsilon'}{2}\left[1 + \left(\frac{60\lambda\sigma}{\varepsilon'}\right)^2\right]^{1/2}}$$

$$\cong \sqrt{-\frac{\varepsilon'}{2} + \frac{\varepsilon'}{2} + \frac{(60\lambda\sigma)^2}{4\varepsilon'}} = 30\lambda\sigma/\sqrt{\varepsilon'} \tag{1.34}$$

where λ is the wavelength in air.

Substituting the values of n and p in the expressions for δ, u, and λ', we get

$$\delta = \frac{\omega}{c} p \cong \frac{60\pi\sigma}{\sqrt{\varepsilon'}} \text{ m}^{-1} \tag{1.35}$$

$$u \cong \frac{c}{\sqrt{\varepsilon'}} = \frac{3 \times 10^8}{\sqrt{\varepsilon'}} \text{ m/s} \tag{1.36}$$

$$\lambda' \cong \frac{\lambda}{\sqrt{\varepsilon'}} \text{ m} \tag{1.37}$$

In cases where the displacement current is only a fraction of the conduction current, that is, where by (1.7)

$$\varepsilon' \ll 60\lambda\sigma$$

Eqs. (1.31) and (1.32) take the form

$$n \cong \sqrt{30\lambda\sigma} \tag{1.38}$$

$$p \cong \sqrt{30\lambda\sigma} \tag{1.39}$$

Substituting the values of n and p in the expressions for δ, u, and λ' we get

$$\delta = \frac{\omega}{c} p \cong 2\pi \sqrt{\frac{30\sigma}{\lambda}} \text{ m}^{-1} \tag{1.40}$$

$$u \cong \frac{3 \times 10^8}{\sqrt{30\lambda\sigma}} \text{ m/s} \tag{1.41}$$

$$\lambda' \cong \frac{\lambda}{\sqrt{30\lambda\sigma}} \text{ m} \tag{1.42}$$

Example 1.2. Find the distance over which the amplitude of the field of a wave propagated in dry soil ($\varepsilon' = 4$, $\sigma = 10^{-3}$ S/m) and in sea water ($\varepsilon' = 80$, $\sigma = 4$ S/m) is reduced to $1/10^6$ of its original value. Find also the velocity of propagation and wavelengths in the two media, taking wavelengths of 20,000, 2000, 200, 20, and 2 metres.

Solution: To begin with, using (1.7), find the ratio j_d/j_c so as t decide which of the three equations (the complete one or any of th two simplified) should be used.

Once the value of δ for each of the given wavelengths is found determine the distance x from $e^{-\delta x} = 10^{-6}$:

$$x = \frac{6 \ln 10}{\delta} = 13.81/\delta \text{ m}$$

The results are summarized in Tables 1.4 and 1.5.

Table 1.4

$\varepsilon' = 4$; $\sigma = 0.001$ S/m

λ, m	j_d/j_c	Design equation	p	δ, m^{-1}	x, m	n	u, m/s	$\lambda' = \lambda/n$,
20 000	3.3×10^{-3}	(1.40)	—	7.66×10^{-3}	1800	24.5	1.22×10^{7}	816
2 000	3.3×10^{-2}	(1.40)	—	2.43×10^{-2}	570	7.75	3.86×10^{7}	258
200	0.33	(1.27)	2.08	6.54×10^{-2}	210	2.88	1.04×10^{8}	69.5
20	3.3	(1.27)	0.30	9.42×10^{-2}	147	2.02	1.34×10^{8}	9.9
2	3.3	(1.35)	—	9.42×10^{-2}	147	2.00	1.5×10^{8}	1.0

Table 1.5

$\varepsilon' = 80$; $\sigma = 4$ S/m

λ, m	j_d/j_c	Design equation	p	δ, m^{-1}	x, m	n	$u=c/n$, m/s	$\lambda'=\lambda/n$, m
20 000	1.67×10^{-5}	(1.40)	—	0.486	28.4	1550	1.93×10^{5}	12.9
2 000	1.67×10^{-4}	(1.40)	—	1.54	8.96	489	6.13×10^{5}	4.1
200	1.67×10^{-3}	(1.40)	—	4.86	2.84	155	1.93×10^{6}	1.3
20	1.67×10^{-2}	(1.40)	—	15.4	0.896	49	6.13×10^{6}	0.43
2	0.167	(1.27)	14.3	45	0.31	17	1.78×10^{7}	0.12

From the above, we may conclude that the attenuation of radio waves drastically increases with frequency, and so the longer waves should be preferred for communication. The attenuation is considerable in magnitude—even on the longest wavelength the amplitude of the field is reduced to one-millionth of its original value over a distance of about 30 metres in sea water and about 1800 metres in dry soil.

Reflection of Radio Waves from the Earth's Surface

When dealing with the propagation of short and ultra-short radio waves we are often concerned with their reflection from the earth's surface. The importance of this phenomenon will be clear from the example that follows.

Let there be a radio wave incident upon the earth's surface at an angle γ near a receiving aerial set up at point A, h metres above the perfectly conducting ground (Fig. 1.13). Let also the wave produce an electric field $E_1 = E_m e^{i\omega t}$ at point A. It would be wrong to think that the emf in the aerial is solely due to this field. The point is that a plane wave incident upon the earth's surface at an angle γ may be resolved in the sum total of parallel rays one of which, *1* in the figure, hits the aerial directly, while another, *2*, does so only after it is reflected from the ground. Hence the total field at the aerial is the sum of the fields due to the direct and reflected rays.

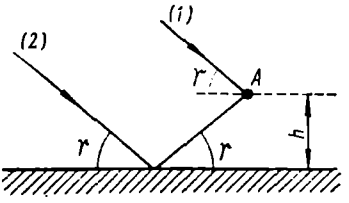


Fig. 1.13. The effect of reflected rays on the resultant field at A

From field theory it is known that reflection from an imperfect conductor changes both the amplitude and the phase of the wave. The amplitude of the reflected wave is always smaller than that of the incident one, while the phase shift may generally be of any value. The change in amplitude and the shift in phase in reflection can be best described in terms of the *complex reflection coefficient*, $\mathbf{R} = Re^{i\theta}$.[10] Setting the instantaneous electric field of the incident wave at the interface as $E_m e^{i\omega t}$ the instantaneous electric field of the reflected wave, also at the interface, will be given by

$$\mathbf{R}E_m e^{i\omega t} = RE_m e^{i(\omega t - \theta)} \tag{1.43}$$

where R is the magnitude of the complex reflection coefficient, and θ its phase. R gives the reduction in amplitude, and θ, the shift (phase) in reflection.

Analysis of wave reflection shows that the magnitude and phase of the complex reflection coefficient depend on the form of polarization of the incident wave. Polarization is defined as the orientation of the electric field vector relative to the direction of propagation and the reference plane, which is usually the earth's surface. Thus, there are vertical and horizontal polarization. Fig. 1.14 shows vertical polarization—the electric field vector is contained in a vertical plane passing through the incident ray (the plane of the drawing). In a ho-

rizontally polarized wave, the electric field vector, which is alway normal to the direction of propagation, is parallel with the earth' surface, as shown by the dot in Fig. 1.15.

In the general case, an incident ray may have a horizontal an a vertical component. If the two components are in phase or ar 180° out of phase with each other, this is a *linearly* or *plane polarize wave*. If the two components have the same peak value but are 90 o 270 degrees out of phase, this is a *circularly* (*right-handed* and *left handed*, respectively) *polarized wave*. When the amplitudes and phase of the two components are related at random, an *elliptically polarize*

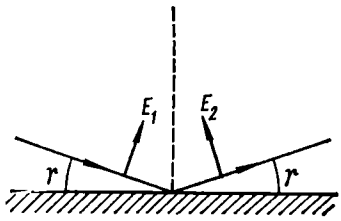


Fig. 1.14. Vertical polarization of an incident wave

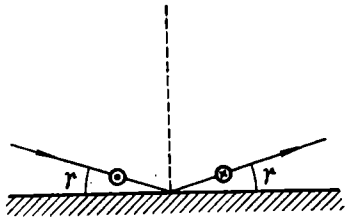


Fig. 1.15. Horizontal polarization of an incident wave

wave results. From knowledge of reflection for vertical and horizonta polarization, it is an easy matter to determine all field component for any polarization of the incident wave.

The problem may be stated as follows. Given: the peak electri field E_m of the incident wave, type of polarization, the grazing angl (the angle between the incident ray and the ground) γ, wavelengt λ, dielectric constant ε' of the ground, and ground conductivity σ To find: R and θ.

With sufficient ease, these quantities can be found from the plot of Fig. 1.16 and of Fig. 1.17 due to Ohman [1, 2]. The former appl to vertical, and the latter, to horizontal polarization.

For vertical polarization, the plot of Fig. 1.16*a* is entered with ε and 60 $\lambda\sigma$, and the values of R_{90}, γ_{00}, and sin γ_{90} are found as th intersections of the respective abscissae and ordinates (if necessary resort may be made to interpolation). The value γ_{90} is defined as th grazing angle for which the shift in phase in reflection is 90 degrees R_{90} is the magnitude of the reflection coefficient at γ_{90}.

Then by the equation

$$\rho = \sin\gamma/\sin\gamma_{90} \qquad (1.44$$

the value of the parameter ρ is found, or its reciprocal, $1/\rho$, when $\rho >$ > 1. Entering the plots of Fig. 1.16*b* and *c* with the values of

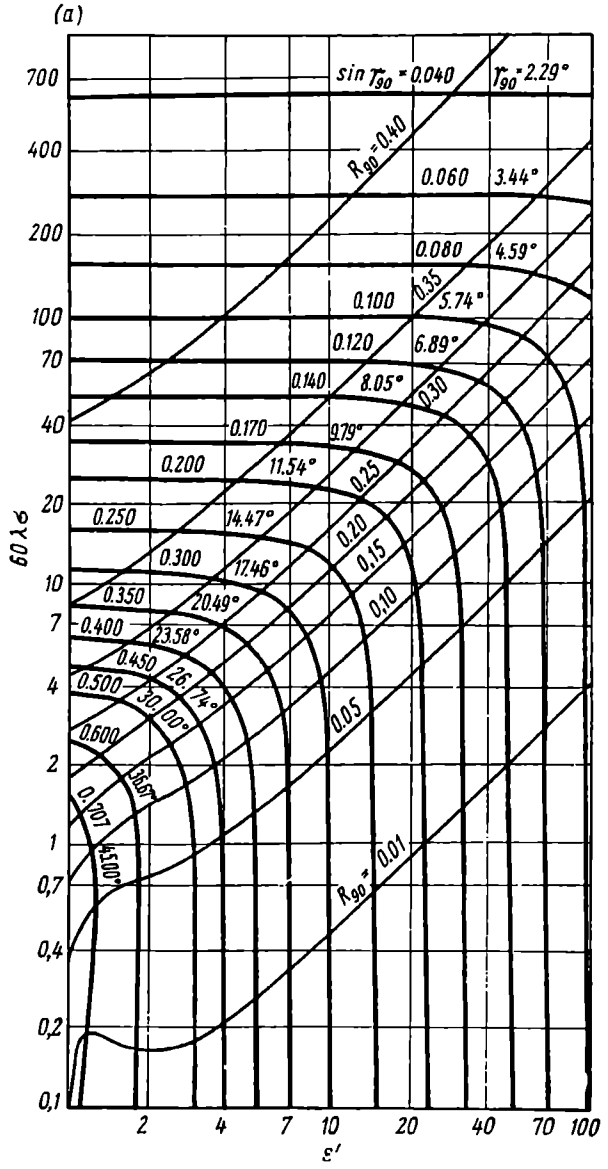


Fig. 1.16*a*. Determination of the reflection coefficient R_{90} and the grazing angle γ_{90} (for a vertically polarized wave)

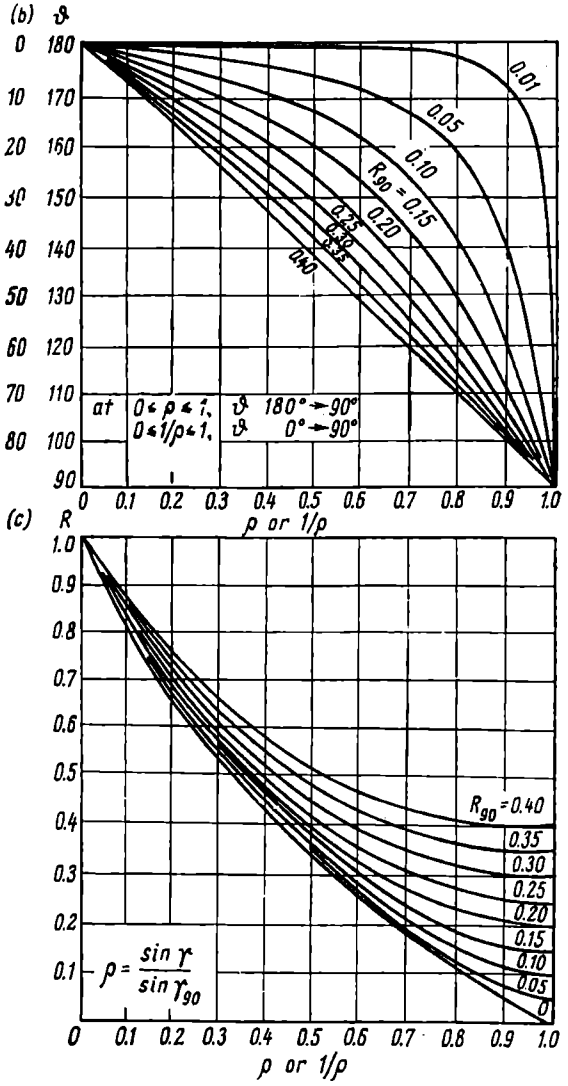


Fig. 1.16b. The phase θ of the reflection coefficient for vertical polarization

Fig. 1.16c. The magnitude R of the reflection coefficient for vertical polarization

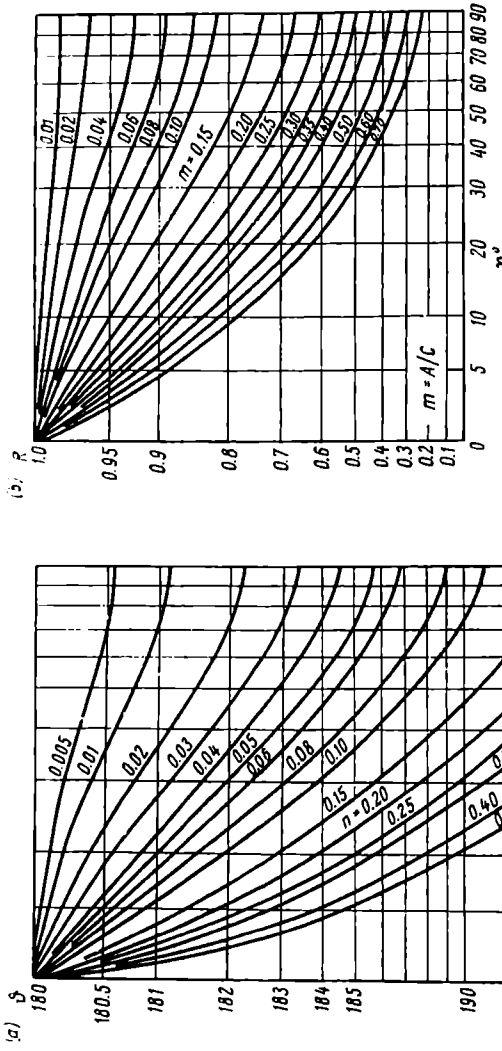


Fig. 1.17. Phase θ (*a*) and magnitude R (*b*) of the reflection coefficient as a function of angle of incidence (for horizontal polarization)

or $1/\rho$ and R_{90}, the phase θ and the magnitude R of the reflection coefficient are found.

The angle γ_{90} is the so-called *pseudo-Brewster angle*, that is, one at which the reflection coefficient of a vertically polarized wave from an imperfectly conducting surface is a minimum. The Brewster angle, or polarizing angle, of a dielectric, is that grazing angle for which a wave polarized parallel to the plane of incidence (vertical polarization) is wholly transmitted (there is no reflection) into another medium as a refracted ray. In Fig. 1.16, the Brewster angle correspond to $R_{90} = 0$.

In the case of pseudo-Brewster grazing angles, the phase shift in reflection is $\theta = 90°$, which is the reason for the subscript "90" in γ_{90} and R_{90}.

We define the quantities

$$C = \sqrt{(\varepsilon' - 1)^2 + (60\lambda\sigma)^2} \tag{1.45}$$

$$A = \sqrt{\frac{C + \varepsilon' - 1}{2}} \tag{1.46}$$

$$B = \sqrt{\frac{C - \varepsilon' + 1}{2}} \tag{1.47}$$

$$m = A/C \tag{1.48}$$

$$n = B/C \tag{1.49}$$

Entering the plot of Fig. 1.17*a* with the given grazing angle and the calculated value of n, the phase θ of the reflection coefficient is found. Similarly, entering the plot of Fig. 1.17*b* with the grazing angle γ and the parameter m, the magnitude of the reflection coefficient is found.

Example 1.3. Using the plots of Figs. 1.16*a*, *b*, and *c*, and 1.17*a* and *b*, determine the phase and magnitude of the reflection coefficient for a damp soil ($\varepsilon' = 10$, $\sigma = 0.01$ S/m), if $\lambda = 6$ m and the grazing angle $\gamma = 10°$ ($\sin \gamma = 0.174$).

Solution:

Vertical polarization: $\varepsilon' = 4$ and $60\lambda\sigma = 3.6$. From the plot of Fig. 1.16*a*, $R_{90} =$ (approx.) 0.01, and $\sin \gamma_{90} = 0.300$. The parameter $1/\rho = 0.174 \div 0.300 = 0.58$. From the plot of Fig. 1.16*b* $\theta = 2°$, and from that of Fig. 1.16*c*, $R = 0.26$.

Horizontal polarization: Find m and n, using Eqs. (1.45) through (1.49). $C = 4.68$; $A = 1.96$; $B = 0.916$; $m = 0.418$; $n = 0.196$. From the plot of Fig. 1.17*a*, $\theta =$ (approx.) 188°, and from that of Fig. 1.17*b*, $R =$ (approx.) 0.093.

1.7. Attenuation Function. Design Principles for Radio Links with Time-invariant Parameters

The propagation of radio waves along the earth's surface is inevitably accompanied by absorption and other losses of energy. Ground waves, for example, lose some of their energy as they penetrate the earth and also due to diffraction (since only a small fraction of the transmitted energy is carried by the diffracted wave around the curvature of the globe). Radio waves propagated through the troposphere mainly pass through it and only a small fraction is scattered to form the useful energy. Ionospheric waves suffer losses as they travel through the imperfectly conducting ionised layers of the atmosphere.

The attenuation of radio waves under real conditions in comparison with free-space propagation may be described in terms of the *attenuation function* F.

This section will deal with simple radio circuits whose parameters are time-invariant and for which F is a fixed quantity.

For such circuits, using the value of F found in some way, the electric field may be expressed as

$$E_{rms} = \frac{173\sqrt{P_{1\text{kW}}G_1}}{r_{\text{km}}} F \text{ mV/m} \tag{1.50}$$

where r is the propagation path length.

Similarly, the Poynting vector is

$$S = \frac{P_1 G_1}{4\pi r^2} F^2 \text{ W/m}^2 \tag{1.51}$$

or in more convenient units,

$$S = 8 \times 10^4 \frac{P_{1\text{kW}}G_1}{r^2_{\text{km}}} F^2 \text{ nW/m}^2 \tag{1.51a}$$

In Eqs. (1.51) and (1.51a), the attenuation function is squared because the Poynting vector is proportional to the square of the field strength, Eq. (1.11).

Allowing the attenuation, the transmission loss, Eq. (1.20), may be expressed as

$$L = P_1/P_2 = L_{fs}/F^2 = \left(\frac{4\pi r}{\lambda F}\right)^2 \frac{1}{G_1 G_2} \tag{1.52}$$

and the basic transmission loss as

$$L_b = L_{bfs}/F^2 = \left(\frac{4\pi r}{\lambda F}\right)^2 \tag{1.53}$$

or, in decibels,

$$L_{\text{db}} = 10\log_{10} L = L_{fs\text{db}} - 20\log_{10} F = 20\log_{10}(4\pi r) - \\ - 20\log_{10}\lambda - F_{\text{db}} - G_{1\text{db}} - G_{2\text{db}} \tag{1.52a}$$

and

$$L_{b,\,\mathrm{db}} = 10\log_{10} L_b = L_{b,\,fs,\,\mathrm{db}} - 20\log_{10} F = \\ = 20\log_{10}(4\pi r) - 20\log_{10}\lambda - F_{\mathrm{db}} \qquad (1.53a)$$

It should be noted that F_{db} for the field strength will have the same absolute value as F^2_{db} for the power flux, that is,

field attenuation: $F_{\mathrm{db}} = 20\log_{10}F$

power-flux attenuation: $(F^2)_{\mathrm{db}} = 10\log_{10}F^2 = 20\log_{10}F$

Therefore, the attenuation function can be simply expressed in decibels, and it is immaterial whether it is in terms of field strength or power flux.

Eq. (1.52) can serve as a convenient tool in the design of radio links. Ordinarily, the specifications for a radio link include the minimum discernible power P_2 (the one at the receiver input required for reliable reception), wavelength λ, operating range r, and attenuation function F decided by the mode of radio propagation, and one is to determine the power to be radiated by the aerial and the gains of the transmitting and receiving aerials. In such cases, Eq. (1.52) may be re-written as

$$P_1 G_1 G_2 = \left(\frac{4\pi r}{\lambda F}\right)^2 P_2 \text{ W} \qquad (1.54)$$

The values of P_1, G_1, and G_2 should be chosen so as to minimize the cost of the radio link.

The problem is simplified when the specification includes the types of aerial and, consequently, the aerial gains.

The problem can be solved logarithmically. For this, the power at the receiver input, P_2, and the transmitted power should be expressed in decibels relative to one watt. Then

$$10\log_{10} P_{2\mathrm{W}} = P_{2\mathrm{dbW}} \text{ and } 10\log_{10} P_{1\mathrm{W}} = P_{1\mathrm{dbW}},$$

where the subscript "dbW" is "decibels relative to one watt".

Eq. (1.54) will then take the form

$$P_{1\mathrm{dbW}} + G_{1\mathrm{db}} + G_{2\mathrm{db}} = P_{2\mathrm{dbW}} + 20\log_{10}(4\pi r) - 20\log_{10}\lambda - F_{\mathrm{db}} \qquad (1.55)$$

or

$$P_{1\mathrm{dbW}} + G_{1\mathrm{db}} + G_{2\mathrm{db}} = P_{2\mathrm{dbW}} + L_{b,\,fs\,\mathrm{db}} - F_{\mathrm{db}} \qquad (1.55a)$$

or

$$P_{1\mathrm{dbW}} = P_{2\mathrm{dbW}} + L_{b,\,\mathrm{db}} - F_{\mathrm{db}} \qquad (1.55b)$$

or, finally,

$$P_{1\mathrm{dbW}} = P_{2\mathrm{dbW}} + L_{\mathrm{db}} \qquad (1.55c)$$

As is seen, the calculation of the radio-link parameters reduces to the operations of addition and subtraction.

From Equations (1.54) and (1.55) it follows that from the standpoint of radio propagation, the crucial problem in the design of radio

links whose parameters are time-invariant is to determine the attenuation function F for the various modes of propagation—ground waves, tropospheric waves, and ionospheric waves. In effect, the remaining sections of the book are devoted to determining the attenuation function for the various cases of radio propagation encountered in practice.

Example 1.4. Find the transmitter power for reliable communication if $P_2 = 10^{-14}$ W or $P_2 = 10\log_{10}10^{-14} = -140$ dbW; $r = 400$ km; $\lambda = 20$ cm; $F = -80$ db; $G_1 = G_2 = 30$ db.

Solution: Inserting the numerical values in Eq. (1.55) gives

$$P_1 = -140-30-30+134+14+80=28 \text{ dbW}$$

Hence, $28 = 10\log_{10}P_1$, and the transmitter power

$$P_1 = 630 \text{ W}$$

In this rather simplified example we have neglected random fluctuations in the attenuation function. Otherwise, the requisite power would be different.

1.8. Signal Distortion in Propagation. Design Principles for Radio Links with Time-varying Parameters

Apart from attenuation, radio waves are subject to distortion. Among other things, the rectangular pulses radiated by transmitters have their waveform distorted at the point of reception due to the dispersion of the medium, that is, the change in the velocity of propagation with frequency. This property is displayed by the earth's imperfectly conducting surface and ionized gases.

Another cause of distortion is the fact that a single transmitted ray acquires additional components along the propagation path, so that the receiving aerial picks up several rays, each covering a different distance, for example, waves reflected from the ionosphere more than once.

The situation is aggravated still more by the fact that, as already noted, the troposphere and the ionosphere are dynamic media whose properties continually vary with time. As a result, the difference in distance covered by the various rays also varies continually. This results in *fading*, that is, random variations in the amplitude and phase of the received signal, and sometimes in echoes.

Sometimes, the areas that reflect or scatter radio waves move in space, causing the Doppler shift in the frequency of the received signal, which likewise results in distortion. In other cases, distortion is caused by nonlinear phenomena in the atmosphere, its anisotropy, impulse noise, or interfering stations. Last but not least, signals are

always received against the background of atmospheric, cosmic, and internal (receiver) noise.

In the case of fading, the attenuation function F should be treated as a random time function. However, the distribution of this random variable is a single-valued function of a particular mode of propagation. Once this distribution is known, it is an easy matter to determine the probability of F exceeding a specified value.

Where distortion manifests itself as a change in the signal waveform or as echo signals, the communication channel can be treated as an equivalent four-terminal network of definite frequency and phase responses.

For reliable operation, a communication channel should simultaneously meet the following two requirements: (1) it should have a specified signal-to-noise ratio (to ensure the reliability and form of operation expected); and (2) the distortion produced along the propagation path should remain within the specified limits (which are decided by the form of operation and the quality of the communication channel desired).

The design procedure for radio links with randomly varying parameters is the same as outlined in § 1.7, except that, as already noted, the attenuation function is a random variable having a definite distribution. With such an approach, using a generalised form of Eq. (1.54), it is possible to determine the value of the product $P_1G_1G_2$ such that the communication channel will have the specified reliability. Worked examples of this procedure will be found in the succeeding sections of the text.

Thus, the objectives of the text are: (1) to teach the reader to determine the attenuation function and its distribution with time under a multitude of propagation conditions; (2) to furnish knowledge of the mechanisms by which the various forms of signal distortion come about; and (3) to outline methods for distortion control.

Chapter Two

Ground-Wave Propagation

2.1. Electrical Properties of the Earth's Surface

Ground waves are propagated close to the earth's surface and, therefore, suffer varying amounts of attenuation according to its nature—surface irregularities, permittivity, and conductivity.

Seventy-one per cent of the globe's surface is covered with water which is different from land in electrical properties. But even water is not always the same in that respect—the saline water of seas and oceans is unlike the fresh water of lakes and rivers. With no wind blowing, a lake or river will present a mirror-smooth surface, while seas and oceans are mostly rippled or rough—sometimes sea waves rise tens of metres high. Even a frozen sea is often covered with ice folds that high.

Land, too, has widely differing kinds of surface—damp soil as in fields and meadows, dry soil as in deserts, plains covered with bushes and forests, and barren rocks and mountains.

As to surface irregularities, their effect on radio propagation varies with wavelength. For very long (VLF) and long (LF) waves running into tens of kilometres, all kinds of the earth's surface, except mountains, are smooth. In contrast, decimetric (UHF) and centimetric (SHF) waves will have a rough going over low sea waves or vegetation.

In studies of ground-wave propagation, one has to idealise the actual conditions. For example, it is impossible to keep track of all continuous variations in surface along the propagation path. Instead, notice is taken of only sudden changes, such as the coast-line, the boundary between a field and a large forest, etc. Nor is it possible to follow variations in the electrical properties of surface with depth (for example, the conductivity of ground increases as it approaches ground waters). Therefore, for design purposes, a real surface non-uniform with depth is replaced with an equivalent, uniform one which will cause the same amount of attenuation.

All kinds of the earth's surface may be divided into two broad groups: (1) surfaces with insignificant irregularities, which may be

replaced with a plain surface having the same electrical characteristics; and (2) surfaces with marked irregularities, which must be replaced with a plain surface having *equivalent electrical characteristics.*

The equivalent electrical characteristics are chosen such that the equivalent plain surface will produce the same amount of attenuation in a specified band of frequencies as the actual one.

Typical examples of surface belonging to the second group are forests and built-up areas. With waves propagated over a forest the attenuation is mainly due to the current they induce in the trunks and branches of trees which may be treated as a sort of earthed aerials made from an imperfect conductor. Of course, the energy of radio waves is absorbed in a forest by a mechanism which has nothing in common with the absorption by a smooth, imperfectly conducting surface. Yet the electrical characteristics of the equivalent (imaginary) surface can be chosen such that its effect in terms of wave attenuation will be the same as that due to the forest being approximated. The same technique applies to radio propagation over built-up areas.

A summary of actual electrical characteristics for surfaces in the first group and equivalent characteristics for various kinds of surfaces is given in Table 2.1.

Calculations of ground-wave propagation call for accurate knowledge of the electrical characteristics of surfaces all over a country. For this purpose, most countries carry on large-scale work to compile surface-conductivity maps. In the Soviet Union, much help in this respect comes from radio amateurs. To assist them, popular-science magazines have published detailed instructions on what to do and how.

Table 2.1. ELECTRICAL CHARACTERISTICS OF EARTH'S SURFACES

Surface	Range of values		Average values	
	ε'	σ, S/m	ε'	σ, S/m
Sea water	80	1-4.3	80	4
Fresh water of lakes and rivers	80	10^{-3}-2.4×10^{-2}	80	10^{-3}
Moist soil	10-30	3×10^{-3}-3×10^{-2}	10	10^{-2}
Dry soil	3-6	1.1×10^{-5}-2×10^{-3}	4	10^{-3}
Forest*	—	—	—	10^{-3}
Built-up areas*	—	—	—	7.5×10^{-4}
Mountains	—	—	—	7.5×10^{-4}

* These are equivalent values for the medium-wave (MF) band.

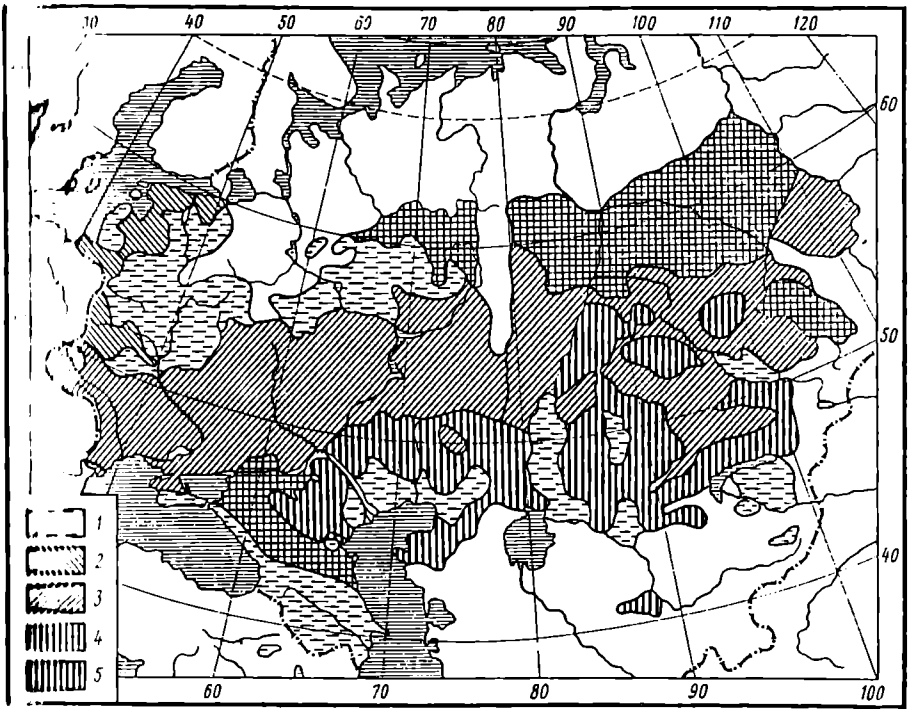


Fig. 2.1. A soil-conductivity map of the Soviet Union (1) $\sigma = 15$ to 20×10^{-3} S/m; (2) $\sigma = 30$ to 40×10^{-3} S/m; (3) $\sigma = 45 \times 10^{-3}$ S/m; (4) $\sigma = 65 \times 10^{-3}$ S/m; (5) $\sigma = 70$ to 90×10^{-3} S/m

Fig. 2.1. shows a surface-conductivity map of the Soviet Union, compiled by Kashparovsky [3], covering the European USSR, the Trans-Ural area, and Soviet Central Asia.

2.2. Ground-wave Propagation over a Flat Earth with Elevated Transmitting and Receiving Aerials

It is natural to begin a study into ground-wave propagation with the simplest case—the *flat-earth problem.*

When the transmitting and receiving aerials are only a short distance apart, it is safe to ignore the curvature of the earth's surface and to think that radio waves are propagated along a flat imperfectly conducting surface. Furthermore, it may be assumed that the surface is smooth and uniform the entire length of the propagation path. This is a fairly common occurrence in practice.

Before going any further, let us define the term "elevated aerial". As applied to ground-based transmitters or receivers, this term implies that: (a) an aerial uses a non-radiating feeder (transmission line), as shown in Fig. 2.2; in other words, a very tall aerial with a radiating down-lead will not be classed as elevated (Fig. 2.3); (b) the aerial height, h, must be several times the wavelength.

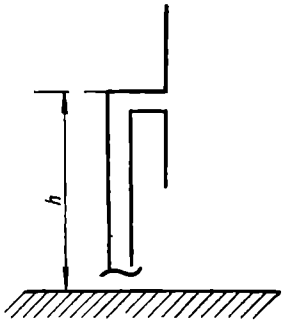


Fig. 2.2. Explaining the concept of an elevated aerial

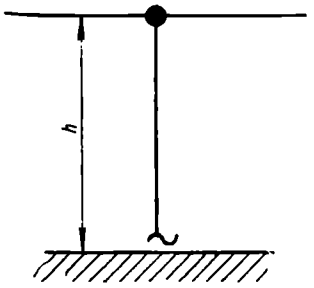


Fig. 2.3. A tall aerial which is not "elevated"

Practically, ground-based transmitters use elevated aerials only on short and ultra-short waves. Typical elevated aerials are those used by TV centres, VHF-FM transmitters, multiple-link relay systems, and radar. Naturally, air-borne radio equipments use elevated aerials on nearly all frequency bands.

With elevated transmitting and receiving aerials, the rigorous solution of the attenuation-function problem, which reduces to solving Maxwell's equations for specified boundary conditions, may be markedly simplified, and the received electric field may be found by the methods of geometrical optics. The problem may be stated as follows.

Given: the geometrical characteristics, such as the distance r between the aerials; the height of the transmitting aerial, h_1, the height of the receiving aerial, h_2; and the electrical characteristics, such as the transmitted power P_1, the gains G_1 and G_2 of the transmitting and receiving aerials, respectively; wavelength λ; type of polarization; the dielectric constant ε' of ground; ground conductivity σ.

To find: the attenuation function F, that is, to describe the received field as

$$E_{rms} = \frac{173\sqrt{P_{1\text{kW}}G_1}}{r_{\text{km}}} F \text{ mV/m} \qquad (2.1)$$

which is the same, to describe the received power as

$$P_2 = \frac{6.33 \times 10^3 P_{1\text{kW}} G_1 G_2 \lambda_m^2}{r_{\text{km}}^2} F^2 \text{ nW} \tag{2.2}$$

As was shown by Vvedensky for the first time in 1922 [4] the electric field of a radio wave at the point of reception may be treated as the

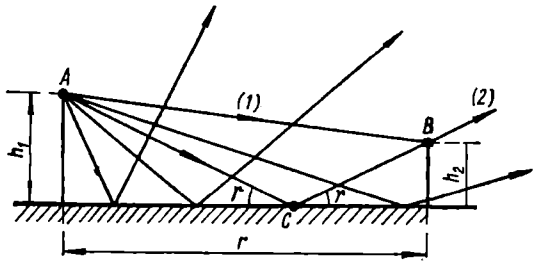


Fig. 2.4. The wave field at the receiving aerial (case 1)

sum of a direct ray *1* (Fig. 2.4) and a ground-reflected ray *2* picked up by the receiving aerial. It should be noted that from among the multitude of rays reflected from the ground, only one (as shown in Fig. 2.4) strikes the aerial.

Then the resultant field may be expressed as the sum of two instantaneous fields: one due to the direct ray

$$E_1 = \frac{245 \sqrt{P_{1\text{kW}} G_1}}{r_{\text{km}}} e^{i\omega t} \text{ mV/m} \tag{2.3}$$

and the other due to the ground-reflected ray

$$E_2 = \mathbf{R} \frac{245 \sqrt{P_{1\text{kW}} G_1}}{r_{2\text{km}}} e^{i\left(\omega t - \frac{2\pi}{\lambda} \Delta r\right)} \text{ mV/m} \tag{2.4}$$

Equations (2.3) and (2.4) need some elaboration.

For one thing, our further discussion will be limited to cases where the aerial height is a small fraction of the horizontal distance, such that $h_1 \ll r$ and $h_2 \ll r$. That is, we shall only deal with ground paths. Because of this, the vertical scale in Figs. 2.4 and 2.5 has had to be exaggerated, and the angles shown are not the true ones. Actually, the angles γ and ψ are very small and measure fractions of a degree. For another, rays *1* and *2* travel away from the transmitting aerial in about the same direction, which is the reason why the gain in Eqs. (2.3) and (2.4) is the same for both aerials.

In Eqs. (2.3) and (2.4), r_1 stands for the path length AB of the direct ray, and $r_2 = AC + CB$ for the path length of the reflected ray. Hence, $\Delta r = AC + CB - AB$ is the difference in path length between the reflected and direct rays. $R = \mathrm{Re}^{-i\theta}$ is the complex reflection coefficient of the ground.

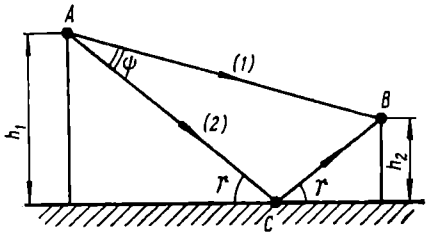


Fig. 2.5. The wave field at the receiving aerial (case 2)

Finally, Eq. (2.4) takes into account the fact that since the path length of the reflected ray is longer by Δr, there appears a phase shift $(2\pi/\lambda)\,\Delta r$. This expression results from the fact that adding a wavelength to the path length shifts the phase through 2π radians.

Since $h_1 \ll r$ and $h_2 \ll r$, the path length of the direct ray differs only slightly from the horizontal distance r. Noting this and inserting the expression for $\mathbf{R}$ in (2.4) gives

$$E_1 = \frac{245\sqrt{P_{1\mathrm{kW}}G_1}}{r_{\mathrm{km}}}\,e^{i\omega t}\ \mathrm{mV/m} \qquad (2.3a)$$

$$E_2 = \frac{R245\sqrt{P_{1\mathrm{kW}}G_1}}{(r+\Delta r)_{\mathrm{km}}}\,e^{i\left(\omega t-\theta-\frac{2\pi}{\lambda}\Delta r\right)}\ \mathrm{mV/m} \qquad (2.4a)$$

From comparison of (2.3a) and (2.4a) it follows that the reflected ray differs from the direct one both in amplitude and phase. The difference in amplitude is caused by the inevitable losses of energy in reflection. The difference in phase is due to two causes: (1) the phase shift θ in reflection; and (2) retardation in phase due to the difference in path length.

In all practical cases $\Delta r \ll r$, so that Δr may be neglected in comparison with r in the denominator of Eq. (2.4a). However, under no circumstances may Δr be neglected in comparison with λ in the exponent.

In view of this, the resultant field may be described as

$$E = E_1 + E_2 = \frac{245\sqrt{P_{1\mathrm{kW}}G_1}}{r_{\mathrm{km}}}\left[1 + Re^{-i\beta}\right]e^{i\omega t}\ \mathrm{mV/m} \qquad (2.5)$$

where $\beta = \theta + (2\pi/\lambda)\,\Delta r$.

After simple manipulations, the expression in the square brackets may be written as the product of the magnitude and the phase factor

$$1 + Re^{-i\beta} = 1 + R\cos\beta - iR\sin\beta = \sqrt{1 + 2R\cos\beta + R^2}\,e^{-i\varphi} \qquad (2.6)$$

where φ is found from

$$\tan\varphi = \frac{R\sin\beta}{1 + R\cos\beta} \qquad (2.7)$$

Recalling that $\beta = \theta + (2\pi/\lambda)\Delta r$ and substituting (2.6) in (2.5), we get

$$E = \frac{173\sqrt{P_{1\text{kW}}G_1}\sqrt{1+2R\cos\left(\theta+\frac{2\pi}{\lambda}\Delta r\right)+R^2}}{r_{\text{km}}} e^{i(\omega t-\varphi)} \text{ mV/m} \quad (2.8)$$

We inquire about the rms value of the resultant field

$$E_{\text{rms}} = \frac{173\sqrt{P_{1\text{kW}}G_1}}{r_{\text{km}}}\sqrt{1+2R\cos\left(\theta+\frac{2\pi}{\lambda}\Delta r\right)+R^2} \text{ mV/m} \quad (2.8a)$$

Comparing it with (2.1), we can readily notice that the attenuation function takes the form

$$F = \sqrt{1+2R\cos\left(\theta+\frac{2\pi}{\lambda}\Delta r\right)+R^2} \quad (2.9)$$

Eq. (2.9) contains three unknowns: the magnitude R and the phase θ of the reflection coefficient, and the path-length difference. From § 1.6 it is known that to find R it is necessary to know the grazing angle γ, type of polarization, and the electric constants of the reflecting surface. Except the grazing angle γ, all the quantities are given in the statement of the problem. Thus, we are to determine the grazing angle γ and the path-length difference Δr.

Referring to Fig. 2.6, the line BC is drawn to cut the extension to the perpendicular dropped from point A onto the ground. Let the point of intersection be A'. The triangle $A'CD$ is equal to the triangle ACD; therefore $AC = A'C$, and $AD = A'D$.

On drawing a straight line parallel with the earth's surface from point B and marking the point of intersection between that line and line AA' as B' we get from the triangle $B'BA'$

$$\tan\gamma = \frac{h_1+h_2}{r} \quad (2.10)$$

which defines the first of the two unknowns.

In most practical cases, $\tan\gamma$ may be replaced with γ; therefore,

$$\gamma \cong \frac{h_1+h_2}{r} \quad (2.11)$$

From the triangles ABB' and $B'BA'$ in the same drawing we get

$$r_1 = AB = \sqrt{r^2+(h_1-h_2)^2} = r\left[1+\left(\frac{h_1-h_2}{r}\right)^2\right]^{1/2} \cong$$

$$\cong r + \frac{h_1^2-2h_1h_2+h_2^2}{2r} \text{ m} \quad (2.12)$$

$$r_2 = A'B = \sqrt{r^2+(h_1+h_2)^2} = r\left[1+\left(\frac{h_1+h_2}{r}\right)^2\right]^{1/2} \cong$$

$$\cong r + \frac{h_1^2+2h_1h_2+h_2^2}{2r} \text{ m} \quad (2.13)$$

On subtracting, we have

$$\Delta r = r_2 - r_1 \cong \frac{2h_1h_2}{r} \text{ m} \qquad (2.1$$

So far, we have neglected the polarization of the transmitted wave. Strictly speaking, Eqs. (2.8) and (2.9) hold solely for a horizontally polarized wave, because with horizontal polarization the vectors of the direct and reflected waves are in the same direction. With vertical polarization, as may be seen from Fig. 2.7, the vecto

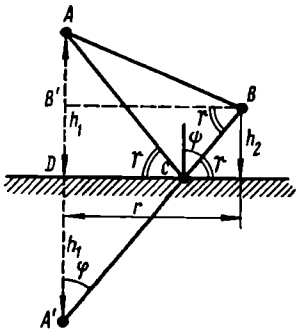


Fig. 2.6. Angle of elevation γ and path-length difference Δr

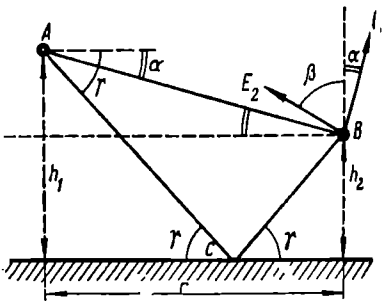


Fig. 2.7. The total field at the receiving aerial for vertical polarization

$\mathbf{E}_1$ is in the plane of the drawing and normal to the line AB, while the vector $\mathbf{E}_2$, which is in the same plane, is normal to CB. However, the angles α and β that these vectors make with the vertical are so small that the vectors $\mathbf{E}_1$ and $\mathbf{E}_2$ may be taken as coincident in direction. If necessary, it will be an easy matter to determine projections of $\mathbf{E}_1$ and $\mathbf{E}_2$ onto the vertical and horizontal (longitudinal) axes, noting that $\alpha = (h_1 - h_2)/r$, and $\beta = \gamma$.

Thus, the resultant field is horizontally polarized in the case of horizontal polarization, and vertically polarized in the case of vertical polarization. Accordingly, when finding the attenuation function from plots of R and θ use should be made of Fig. 1.16 for vertical polarization, and of Fig. 1.17, for horizontal.

As follows from (2.9), the attenuation function F is dependent on the distance r implicitly. To establish the dependence, we shall substitute the expression for Δr from (2.14) in (2.9). Then, the expression for the attenuation function will take the form

$$F = \sqrt{1 + 2R\cos\left(\theta + \frac{4\pi h_1 h_2}{\lambda r}\right) + R^2} \qquad (2.9a$$

It should be remembered that the magnitude (R) and phase (θ) of the reflection coefficient are a function of the reflection point, that is the distance r, because the grazing angle γ varies with r.

As r varies, the value of F passes through a number of maxima, when $\cos\left(\theta + \frac{4\pi h_1 h_2}{\lambda r}\right) = +1$, and a number of minima, when $\cos\left(\theta + \frac{4\pi h_1 h_2}{\lambda r}\right) = -1$. At maxima, $F = (1 + R)$, and at minima, $F = (1 - R)$. Thus, F may be called "attenuation function" only arbitrarily, because at maxima it is greater than unity.

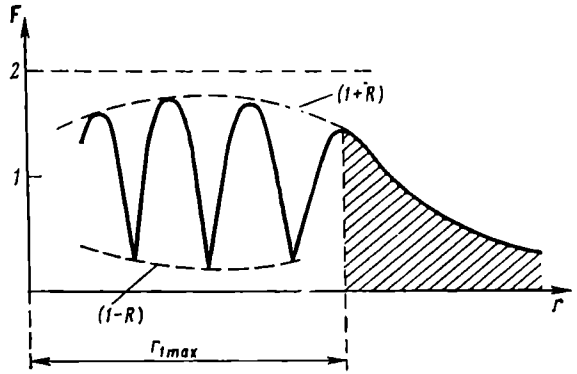


Fig. 2.8. Attenuation function versus distance (the general case)

Typical variations in F as a function of r are shown in Fig. 2.8. The upper broken line is the envelope of maxima, and the lower broken line, that of minima. At some distance r_{1max}, $\theta + \frac{4\pi h_1 h_2}{\lambda r} = 2\pi$. This is the last maximum of the attenuation function, or the first maximum, if we count from the large-distance end. As r increases further, the argument of the cosine will swing between 2π and π (as will be shown shortly, $\theta + \frac{4\pi h_1 h_2}{\lambda r}$ tends towards π at small grazing angles), and F will monotonically decrease tending to zero in the limit. The area of monotonic decrease of the attenuation function is shown shaded in Fig. 2.8.

The reader will probably have noted that the F-vs.-r curve has smooth maxima which look like the tops of a sinusoid, while the minima appear as abrupt changes.

Example 2.1. Find the attenuation function, field at the receiver, and transmission loss for $P_1 = 15$ W, $\lambda = 35$ cm, $G_1 = 100$ (20 db), $G_2 = 100$ (20 db); $h_1 = 80$ m; $h_2 = 20$ m, and $r = 8$ km. The transmission is over a dry ground ($\varepsilon' = 4$; $\sigma = 0.001$ S/m). Consider vertical and horizontal polarization.

Solution: Using (2.11), find the grazing angle

$$\gamma = 100 \div 8000 = 0.08 \text{ radian}$$

By (2.14), $\Delta r = 2 \times 80 \times 20 \div 8000 = 0.4$ m.

Entering the plots of Figs. 1.16 and 1.17 with $\varepsilon' = 4$, $\sigma = 0.001$ S/m, $\gamma = 0.08$, and $\lambda = 35$ cm, find R and θ. For horizontal polarization, $R = 0.91$ and $\theta = 180°$, for vertical polarization, $R = 0.66$ and $\theta = 180°$.

Inserting the values thus found in (2.9) gives:

—for horizontal polarization

$$F = 0.74 \text{ or } F = 20 \log_{10} 0.74 = -2.6 \text{ db}$$

—for vertical polarization

$$F = 0.712 \text{ or } F = 20 \log_{10} 0.712 = -2.9 \text{ db}$$

By Eq. (2.8a), the respective field strengths are:

—with horizontal polarization

$$E_{rms} = \frac{173 \sqrt{15 \times 10^{-3} \times 100}}{8} \times 0.74 = 19.6 \text{ mV/m}$$

—with vertical polarization

$$E_{rms} = \frac{173 \sqrt{15 \times 10^{-3} \times 100}}{8} \times 0.712 = 18.9 \text{ mV/m}$$

Using (1.52a), the transmission loss in decibels is:

—with horizontal polarization

$$L_{db} = 100 + 9.1 + 2.6 - 20 - 20 = 71.7 \text{ db}$$

—with vertical polarization

$$L_{db} = 100 + 9.1 + 2.9 - 20 - 20 = 72.0 \text{ db}$$

In many cases, Eq. (2.9a) may be simplified still more. As follows from the plots of Figs. 1.16 and 1.17, at small grazing angles for most practical types of surface R is unity very nearly and θ is close on 180°. As will be shown later, this is especially true of horizontally polarized waves.

Putting $R \approx 1$ and $\theta = 180°$ in (2.9) and (2.9a), simple manipulations yield

$$F = 2 \left| \sin \frac{\pi}{\lambda} \Delta r \right| = 2 \left| \sin \frac{2\pi h_1 h_2}{\lambda r} \right| \tag{2.15}$$

Expressing r in kilometres and the angle in degrees, we get

$$F = 2 \left| \sin \left(\frac{0.360 h_{1m} h_{2m}}{r_{km} \lambda_m} \right)^\circ \right| \tag{2.15a}$$

In Eqs. (2.15) and (2.15a), the absolute value of the sine is taken, because the minus sign in the radical of (2.9) has had to be dropped in manipulations.

Substituting the expression for F in (2.1) yields an expression for the received electric field

$$E_{rms} = \frac{346\sqrt{P_{1kW}G_1}}{r_{km}} \left| \sin\left(\frac{0.360h_{1m}h_{2m}}{r_{km}\lambda_m}\right)^{\circ} \right| \text{ mV/m} \qquad (2.16)$$

Equations (2.15) and (2.15a) also describe the field as a structure produced by wave interference, because as the distance increases the attenuation function (and, consequently, the field strength) passes through a number of maxima and minima.

The distances of maxima can be found from

$$\frac{2\pi h_1 h_2}{\lambda r} = \frac{\pi}{2}(2n+1) \qquad (2.17)$$

where n may be any integer (0, 1, 2, etc.). Hence,

$$r = \frac{4h_1h_2}{\lambda(2n+1)} \text{ m} \qquad (2.17a)$$

Thus, the first maximum, as counted from the large-distance end, will be

$$r_1 = \frac{4h_1h_2}{\lambda} \text{ m}$$

distant from the transmitter.

The distance of minima can be found from

$$\frac{2\pi h_1 h_2}{\lambda r} = \pi(1+n) \qquad (2.18)$$

where n can take on the same values. Hence

$$r = \frac{2h_1h_2}{\lambda(1+n)} \text{ m} \qquad (2.18a)$$

Thus, the first minimum will be $r_1 = 2h_1h_2/r$ metres distant from the transmitter.

As distinct from the more general case, the maximum value of the attenuation function is two, and the minimum value, zero. This is natural, because no energy is now lost in reflection. A plot of F as a function of r is shown in Fig. 2.9 where the shaded area, as in Fig. 2.8, represents the monotonic decrease in F.

Let us establish in greater detail the conditions under which the simplified equation for the attenuation function holds, that is, the maximum values of $\gamma \cong (h_1 + h_2)/r$ at which we may put $R \cong 1$ and $\theta \cong 180°$ with an accuracy sufficient for practical purposes.

We shall limit ourselves to three types of surface: sea water ($\varepsilon' = 80$, $\sigma = 4$ siemens per metre), moist soil ($\varepsilon' = 10$, $\sigma = 0.01$ siemens per metre), and dry soil ($\varepsilon' = 4$, $\sigma = 10^{-3}$ siemens per metre).

From the plots of Figs. 1.16 and 1.17 it follows that for vertical polarization, operation on shorter waves favours the use of Eq. (2.15). Conversely, with horizontal polarization, the use of Eq. (2.15) is preferable in operation on longer waves.

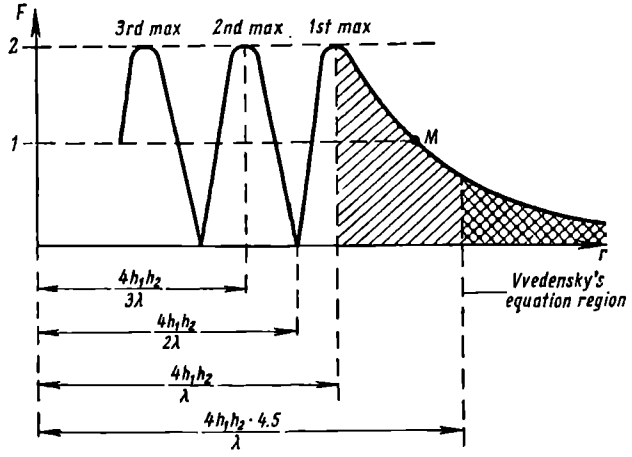


Fig. 2.9. Attenuation function versus distance, with $R = 1$ (the scale of abscissae is exaggerated)

The maximum values of the ratio $(h_1 - h_2)/r$, at which the inequalities $0.95 < R < 1$ and $175° < \theta < 180°$ hold, are summarized in Table 2.2.

Table 2.2. MAXIMUM VALUES OF $(h_1 + h_2)/r$ AT WHICH $R \cong 1$ AND $\theta \cong 180°$

Wavelength λ	Maximum value of $(h_1 + h_2)r$						Polarization
	sea water		moist soil		dry soil		
	$R \cong 1$	$\theta \cong 180°$	$R \cong 1$	$\theta \cong 180°$	$R \cong 1$	$\theta \cong 180°$	
10 m	8×10^{-4}	10^{-3}	8×10^{-3}	6×10^{-2}	10^{-2}	0.25	Vertical
1 cm	3×10^{-3}	0.1	8×10^{-3}	0.3	10^{-2}	0.45	Vertical
10 m	Any	Any	0.08	Any	0.045	Any	Horizontal
1 cm	0.25	Any	0.08	Any	0.045	Any	Horizontal

A study of Table 2.2 displays the following.

For one thing, the conditions $R \cong 1$ and $\theta = 180°$ are satisfied more easily with a horizontally polarized wave. For another, the condition $\theta = 180°$ is much easier to satisfy in all cases than the condition $R \cong 1$. Consequently, the criterion for the validity of

... in the condition $R \cong 1$. Finally, with vertical polarization ... the validity of (2.15) decrease of ground conductivity is favourable ... with horizontal polarization the same is true of surfaces ... the highest conductivity. The worst case is radio propagation over ... water at a wavelength of 10 metres with vertical polarization. Under these conditions, Eq. (2.15) may be used only when ... $h_2)/r < 10^{-3}$, that is, with either very low aerials or with very ... distances. With horizontal polarization, Eq. (2.15) may be ... in practically all cases.

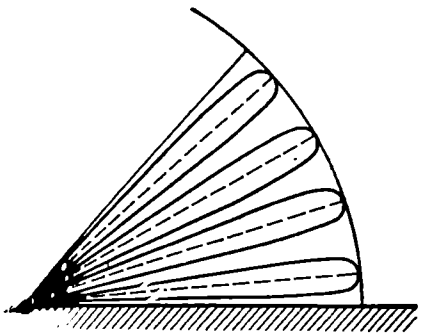


Fig. 2.10. The multi-lobe radiation pattern of a radiator in a vertical plane

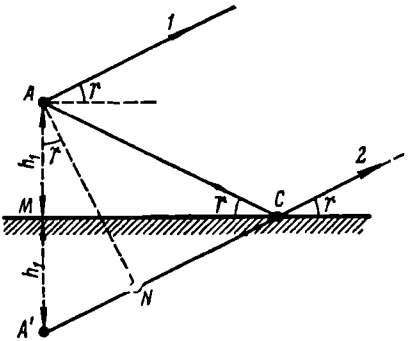


Fig. 2.11. Determining the number of lobes in the radiation pattern of an aerial

Equations (2.9) and (2.15) may be called the (complete and simplified) interference equations because the received field is due to the combination of the direct and reflected rays.

So far we have been concerned with the interference structure of the attenuation function in the direction of the propagation path, that is, with r being varied and h_1 and h_2 held constant. It is as interesting to study the spatial interference structure of the field due to a radiator. At first glance, it may look surprising that an isotropic radiator (that is, one having a circular polar diagram in free space), when placed a distance h_1 above the flat ground, should produce a multi-lobed vertical radiation pattern, as shown in Fig. 2.10. Physically, this is easy to explain because in each direction decided by the grazing angle, two rays, direct *1* and reflected *2*, combine. In some directions, they will be in phase and add together to produce maxima, while in other directions they will be in anti-phase and cancel to produce minima (Fig. 2.11).

Thus, the problem is to establish the dependence of the attenuation function, Eq. (2.15), on the angle of elevation γ within the first quadrant, that is, between zero and 90°, with h_1 and r held constant. If

one wishes so, the receiving aerial may be visualized as tracing o a circle of radius r about a fixed transmitting aerial. The reasoni will not lose any of its generality if we assume that r is large enou for the direct and reflected rays to be taken parallel to each oth (Fig. 2.11).

The attenuation function in the direction of γ can be obtained l use of (2.15). However, (2.14), giving the path-length difference, no longer applicable because it has been derived on the assum tion that $h_1 \ll r$ and $h_2 \ll r$. The second inequality is no longer v lid, and a new expression ought to be derived for the path-leng difference.

Referring to Fig. 2.11, let ray *2* be extended until it cuts an exte sion to the perpendicular dropped onto the surface from point The point of intersection will be A'. The triangles ACM and MC will be equal; consequently, $AC = A'C$, and point A' may l regarded as an imaginary source of the reflected wave. Droppi a perpendicular from A onto the ray $A'C$, we will find that t segment $A'N$ represents the sought difference in path length.

From the triangle $A'AN$

$$\Delta r = 2h_1 \sin \gamma \qquad (2.1$$

Substituting (2.19) in (2.8) gives

$$F = 2 \left| \sin \left(\pi \frac{2h_1}{\lambda} \sin \gamma \right) \right| = 2 \left| \sin (n\pi \sin \gamma) \right| \qquad (2.2$$

where $n = 2h_1/\lambda$.

As γ varies from zero to 90°, sin γ varies from zero to unity, an the argument of the external sine, from zero to $n\pi$. Noting that th period of the absolute value of the sine is π, we may conclude tha the external sine will have n maxima. Consequently, the radiatio pattern of an isotropic aerial will contain as many, that is, $2h_1/l$ lobes within a single quadrant.

The maxima and minima are well defined with a horizontall polarized wave, but less so with a vertically polarized wave sinc the reflection coefficient is then less than unity.

The multiplicity of lobes in a vertical plane is of important practi cal significance, especially in communications or high-altitude rada where it is essential to keep minima out of the radiation pattern i the specified direction.

With directional aerials which themselves display direction effects in a vertical plane, the effect of the ground-reflected r manifests itself in that the lobes described by (2.20) will "cut" t original polar diagram, and the effective total radiation patte will then be as shown in Fig. 2.12, where the broken-line figure sho the radiation pattern without ground reflections, and the full-li one, with ground reflections.

It should be noted that with highly directional aerials the effect of ground reflections manifests itself only at such angles of elevation when the intensity of the ray AC incident upon the ground is comparable with that of the direct ray.

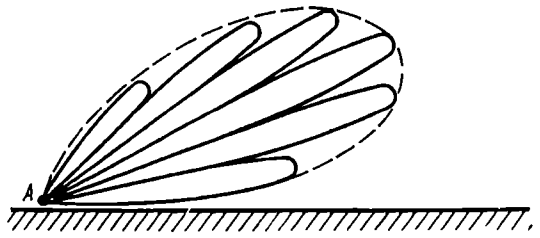


Fig. 2.12. The resultant polar diagram of a radiator in a vertical plane

In many cases of practical importance, Eq. (2.15) can be further simplified. When the argument of the sine is less than about 20 degrees, that is, when

$$\frac{2\pi h_1 h_2}{\lambda r} \leqslant \frac{\pi}{9} \text{ or } h_1 h_2 \leqslant \frac{r\lambda}{18} \text{ m}^2 \tag{2.21}$$

the sine may be replaced with its argument, and the expression for the attenuation function takes the form

$$F = \frac{4\pi h_1 h_2}{\lambda r} \tag{2.22}$$

Inserting it in (2.16) and using more convenient units gives

$$E_{rms} = \frac{2.18\sqrt{P_{1\text{kW}} G_1 h_{1\text{m}} h_{2\text{m}}}}{r_{\text{km}}^2 \lambda_{\text{m}}} \text{ mV/m} \tag{2.23}$$

Equation (2.23) was derived by Vvedensky of the USSR in 1928, and has been named after him [5].

Referring to Fig. 2.9, the region where Vvedensky's equation is valid begins at $r > 18h_1h_2/\lambda$ which is four times the distance to the first maximum. This region is marked by the double shading in Fig. 2.9.

Vvedensky's equation is of fundamental importance to the design of USW (microwave) radio links and gives a clear insight into the dependence of the field on distance, wavelength, and aerial height.

At first glance it may appear unusual that in Eq. (2.23) the total field is inversely proportional to the square of the distance. Incidentally, this is why it is often called "quadratic". The equation also

shows that the total field is inversely proportional to the wavelength used and directly proportional to the product of the aerial heights. In other words, the total field can be enhanced by using a shorter wavelength and increasing the aerial heights, h_1 and h_2.

The cause for this dependence lies in the fact that under the conditions where Vvedensky's equation holds both rays are practically equal in amplitude, but are nearly 180° out of phase. This situation is illustrated by the vector diagram of Fig. 2.13 where E_{m0} is the peak electric field in free space. The same peak field is due to both the direct and reflected rays *1* and *2*. The total phase difference between the rays is the sum of $\theta = \pi$ and $\theta = (2\pi/\lambda)\,\Delta r$, as shown in Fig. 2.13. The resultant vector is nearly normal to the reflected-ray vector. Therefore, the shaded triangle only slightly differs from a right one. Since the angle $(2\pi/\lambda)\,\Delta r$ is very small, we may replace the sine with its argument and get

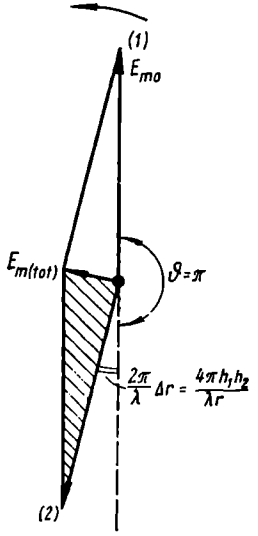


Fig. 2.13. The vector diagram to find the resultant electric field at the point of reception under the conditions of applicability of Vvedensky's equation

$$E_{mr} = E_{m0}\frac{4\pi h_1 h_2}{\lambda r}\ \text{V/m} \qquad (2.24)$$

where E_{mr} is the peak electric field due to the reflected ray. By definition, $4\pi h_1 h_2/\lambda r$ is the attenuation function and it checks with (2.22).

As follows from (2.24), the total field is directly proportional to the complementary angle of the phase shift. This is why the total field is directly proportional to the product of the aerial heights and inversely proportional to the wavelength. Since in free space the field is inversely proportional to the first power of the distance, the overall effect is that the total field is inversely proportional to the square of the distance.

Example 2.2. Determine the attenuation function and the electric field at the point of reception for $P_1 = 50$ W; $\lambda = 10$ cm, $G_1 = 60$; $h_1 = 25$ m, $h_2 = 10$ m and $r = 10$ km. The propagation path is over moist soil. The aerial radiates a vertically polarized wave.

Solution: To begin with, select the equation to use:

$$\frac{h_1 + h_2}{r} = 35 \div 10^4 = 3.5 \times 10^{-3}$$

As follows from Table 2.2, since $3.5 \times 10^{-3} < 8 \times 10^{-3}$, the attenuation function for a vertically polarized wave can be determined by (2.15).

Next, determine whether Vvedensky's equation may be used. Inserting the requisite values in (2.21) gives

$$h_1h_2 = 250 \text{ m}^2; \quad r\lambda/18 = 10^4 \times 0.1 \div 18 = 55 \text{ m}^2$$

that is, (2.21) is not satisfied, and use should be made of (2.15a). Inserting the numerical values in (2.15a), we get

$$F = 2 \left| \sin 90° \right| = 2$$

that is, F is not, in fact, an attenuation function because it is greater than unity.

Using (2.16), find the electric field

$$E_{rms} = \frac{346 \sqrt{50 \times 10^{-3} \times 60}}{10} \left| \sin 90° \right| = 60 \text{ mV/m}$$

Example 2.3. Find the attenuation function and the received electric field for the same variables as in the previous example, using $r = 1$ m.

Solution: Inserting the numerical values in (2.21), we get

$$h_1h_2 = 250 \text{ m}^2; \quad r\lambda/18 = 550 \text{ m}^2$$

that is, (2.21) is satisfied, and use may be made of Vvedensky's equation.

Substituting the numerical values in (2.22) gives

$$F = \frac{4\pi \times 25 \times 10}{1 \times 10^4} = 0.314$$

Substituting the same values in (2.23), we obtain

$$E_{rms} = \frac{2.18 \sqrt{50 \times 10^{-3} \times 60} \times 25 \times 10}{10^2 \times 1} = 9.4 \text{ mV/m}$$

Since under the conditions where Vvedensky's equation holds the reflected ray is detrimental, the parameters of the radio link must be such that the reflected ray will not weaken the direct ray.

Suppose there is a case for which (2.15) is valid. Then, the sought condition will be satisfied if $F = 1$, that is, at

$$2 \left| \sin \frac{2\pi h_1h_2}{\lambda r} \right| = 1,$$

whence

$$\frac{2\pi h_1h_2}{\lambda r} = \pi/6$$

or

$$h_1h_2 = \lambda r/12 \text{ m}^2 \tag{2.25}$$

This condition is satisfied at $r = 3 \times \frac{4\pi h_1h_2}{\lambda}$, represented by point M in Fig. 2.9. This is three times the distance to the first maximum.

Physically, this effect of the reflected ray is due to the fact th it is out of phase with the direct ray by an angle equal to 180° (th phase shift in reflection) plus $(360°/\lambda)\Delta r$ (the phase shift due to th difference in path length).

In our case, from (2.14) and (2.25) it follows that

$$\Delta r = \frac{2h_1h_2}{r} = \lambda/6$$

whence the total phase shift is

$$180° + (360/\lambda)(\lambda/6) = 240°$$

The corresponding vector diagram is shown in Fig. 2.14 from whi it is seen that with a phase difference of 240°, the amplitude of th total field is equal to that of a field i free space.

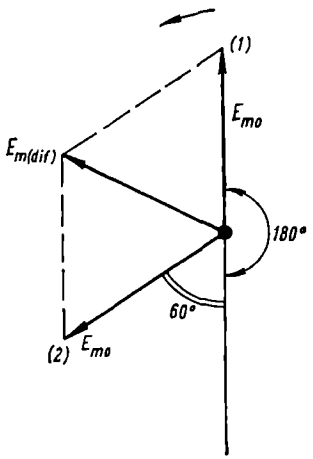


Fig. 2.14. The vector diagram to find the resultant electric field at the point of reception for $\Delta r = \lambda/6$

It should be stressed once more th the concept of direct and reflected ra at the point of reception is valid onl when the aerial height is at least sever times the wavelength. Otherwise, one h to reckon with diffraction, that is, wit the effect of the flat partially conductin surface of the earth on radio propagatio and to use more rigorous methods f determining the received field.

If one of the aerials (no matter which is on the ground, the field cannot b found by the interference equation because they reduce the field to zer when $h_2 = 0$, which is not true. Actually the radiated wave experiences attenuatio due to energy absorption along the entir propagation path, and not only at th point of reflection.

What follows is a simplified desig technique applicable to a fairly wid range of radio-engineering problems involving low aerial heights [0 In this method, the attenuation function is given by

$$F = \frac{4\pi\sqrt{h_1^2+h_0^2}\sqrt{h_2^2+h_0^2}}{\lambda r} \qquad (2.2$$

where h_0 is the *minimum effective aerial height*, which can be foun for sea water, moist and dry soils, and both forms of polarizatio from the plot of Fig. 2.15.

If $h_1 \ll h_0$ and $h_2 \gg h_0$, Eq. (2.26) reduces to (2.22). Conversely, if $h_1 \gg h_0$ and $h_2 \ll h_0$, Eq. (2.26) takes the form

$$F = \frac{4\pi h_0^2}{\lambda r} \tag{2.26a}$$

from which it follows that for low aerial heights the attenuation function is independent of the height of the aerials.

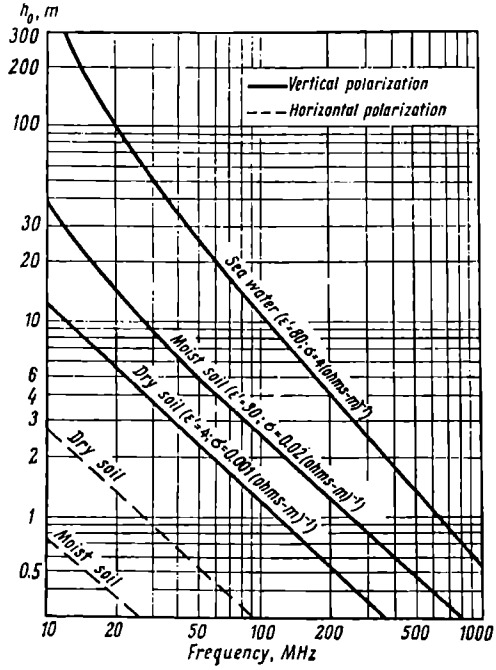


Fig. 2.15. Minimum effective height as a function of frequency for different types of surface

From the plot of Fig. 2.15 it is seen that when a horizontally polarized wave is propagated over sea water, the interference equations may be used for practically any aerial height.

When a soil (or, more generally, any surface) has characteristics other than given in the plot of Fig. 2.15, the minimum effective aerial height may be found as

$$h_0 = \frac{\lambda}{2\pi q} \text{ m} \tag{2.27}$$

where for a vertically polarized wave

$$q = \left| \frac{\sqrt{\varepsilon' - i60\lambda\sigma - \cos^2\gamma}}{\varepsilon' - i60\lambda\sigma} \right| \tag{2.28a}$$

and for a horizontally polarized wave

$$q = |\sqrt{\varepsilon' - i60\lambda\sigma - \cos^2\gamma}| \qquad (2.2)$$

Example 2.4. Find the attenuation function for $\lambda = 1$ m, r 5 km, $h_1 = 6$ m, and $h_2 = 4$ m. The radio wave is vertical polarized and is propagated over a moist soil.

As follows from the plot of Fig. 2.15, for $\lambda = 1$ m ($f = 300$ MH and moist soil, $h_0 = 0.8$ m.

Inserting this and other numerical values in (2.26), we get

$$F = \frac{4\pi\sqrt{6^2+0.8^2}\times\sqrt{4^2+0.8^2}}{1\times5\times10^3} = 0.062$$

By Eq. (2.22), the attenuation function is

$$F = \frac{4\pi6\times4}{1\times5\times10^3} = 0.06$$

2.3. Leontovich's Approximate Boundary Conditions

In rigorous terms, the *flat-earth problem* of radio propagation solved in the following way. To begin with, Maxwell's equations a written for the upper medium (air) and the lower medium (ground Then a solution is sought to these equations, such that the bounda conditions (the conditions at the interface) especially at the radiato and the so-called infinity conditions are satisfied.

In 1944, Leontovich of the Soviet Union showed [7] that the pro lem could be greatly simplified if the exact boundary conditio were replaced with approximate relationships which have come t be known as *Leontovich's conditions*, also known as the *surface impe ance conditions.*

Before going any further, let us recall the exact boundary cond tions of electrodynamics. The first condition is that when an electr magnetic wave passes from one medium to another of different perm ttivity and permeability, that is through a surface of discontinuit the tangential components of the electric and, respectively, th magnetic field are the same on the two sides of the boundary surfac that is, in the two media (the tangential component is said to b continuous across the boundary). The second condition is that th normal components of, respectively, the electric displacement an the magnetic induction are the same on the two sides of the boundary that is, in the two media (the normal displacement and the norma induction are said to be continuous across the boundary).

Deeming that the Z axis is normal to the boundary between th two media (the interface), the exact boundary conditions may b

written as

$$\left.\begin{aligned} E_{1t} &= E_{2t} \\ H_{1t} &= H_{2t} \end{aligned}\right\} z = 0 \tag{2.29}$$

$$\left.\begin{aligned} E_{1z}\varepsilon'_{1c} &= E_{2z}\varepsilon'_{2c} \\ H_{1z}\mu'_1 &= H_{2z}\mu'_2 \end{aligned}\right\} z = 0 \tag{2.30}$$

where the subscript *1* refers to the conditions in the medium on one side of the boundary and the subscript *2* to the conditions in the other medium. The subscript *t* applies to the tangential components.

The idea of the approximate boundary conditions is this. To begin with, they are called approximate because they apply only to cases where the magnitude of the complex relative permittivity is much greater than unity, or

$$\sqrt{(\varepsilon')^2 + (60\lambda\sigma)^2} \gg 1 \tag{2.31}$$

where ε' and σ are the dielectric constant and conductivity of the surface in question. The dielectric constant of air is taken equal to unity.

If (2.31) is satisfied, whether owing to a higher value of ε' or a higher conductivity σ then, as follows from (1.31), the quantity n, which determines the velocity of propagation and, consequently, the wavelength in an imperfectly conducting surface, will also be much greater than unity. Under the circumstances and according to (1.29), the wavelength in the ground

$$\lambda_2 = \lambda_1/n \text{ m} \tag{2.32}$$

will be much shorter than it is in air, λ_1.

Thus, Eq. (2.31) amounts to stating that the wavelength in the ground must be much shorter than it is in air. It should be noted that actual surfaces satisfy Eq. (2.31) nearly always, so that this is not a limitation.

Now let a radio wave of wavelength λ_1 (in air) be propagated in air along the earth's imperfectly conducting surface. The field at point B at some depth h under the interface may be determined by use of Huygens' principle. It recognizes that each point of an advancing wave front is in fact the source of a new train of waves, and that the advancing wave as a whole may be regarded as the resultant of the secondary waves arising from points in the medium (surface) already traversed. Of course, in finding the resultant wave, one should take into account the wave phases at the various points of the medium (surface).

We shall choose the interface between air and ground to be this surface and shall assume, as a first approximation, that the interface is coincident with the equi-phase surface of the advancing wave,

on the wave front (it will be shown shortly that this assumption comes very closely to the actual state of things). In texts on optics (see also § 2.10 of this text) it is shown that to find a field at a point of interest there is no need to carry out summation over the entire surface of the wave front. It will suffice to take notice of the radiators located within the first Fresnel zone.

A Fresnel zone is the circular portion of a wave front transverse to the line between the source and a point of observation, whose

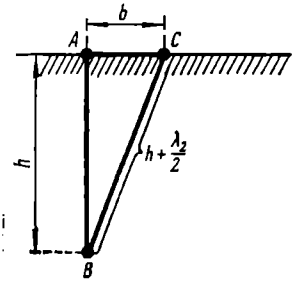


Fig. 2.16. Determination of the wave field at point B under the earth's surface

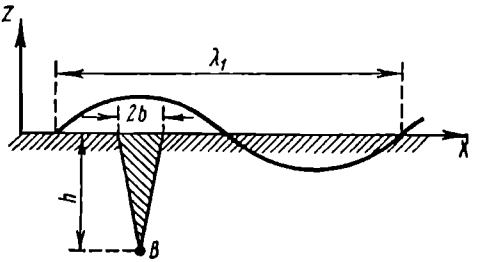


Fig. 2.17. Pertaining to Leontovich's approximate conditions

centre is the intersection of the front with the direct ray and whose radius is such that the shortest path from the source through the periphery to the receiving point is one-half wave longer than the ray. Referring to Fig. 2.16, the shortest path from point B is the depth $h = AB$. If the interface is coincident with an equi-phase surface the first Fresnel zone will be a circle of radius b. The points C on this circle are $BC = h + \lambda_2/2$ distant from the point of observation.

Since

$$h + \lambda_2/2 = \sqrt{h^2 + b^2},$$

the radius b will be given by

$$b = \sqrt{h\lambda_2 + (\lambda_2/2)^2} \text{ m} \tag{2.33}$$

If the depth h at which the field is being determined is much less than the wavelength λ_2, then $b \cong \lambda_2/2$, that is, the first zone will measure about one-half of the wave-length in ground, provided that with the assumptions made $\lambda_2 \ll \lambda_1$.

Let us now refer to Fig. 2.17 which shows a wave propagated in air along the interface in the direction of the X axis. As follows from the foregoing, the dimension $2b$ is much shorter than the wavelength in air, and the field in air over this distance may be regarded as being co-phasal, which justifies the assumption that the interface is an equi-phase surface.

...equently, the field at point B (at a shallow depth) will be ...ed by a small area on the earth's surface, within which the ... may be taken as co-phasal. Such a co-phasal area will generate ... radio waves propagated into the ground in a direction normal ... the interface.

...ield theory recognizes that in a plane radio wave propagated ... an ideal, homogeneous dielectric along the Z axis the magnetic ... electric fields are related as

$$H_y = -\frac{\sqrt{\varepsilon'}}{120\pi} E_x \text{ A/m} \tag{2.34}$$

...sing the concept of the complex permittivity ε_c' introduced ... 1.3 for an imperfectly conducting medium, Eq. (2.34) may be ... written for a plane wave propagated into the earth as

$$H_{2y} = -\frac{\sqrt{\varepsilon_c'}}{120\pi} E_{2x} \text{ A/m} \tag{2.35}$$

...on the strength of the exact boundary conditions expressed by (2.29)

$$H_{2y} = H_{1y}, \quad \text{and} \quad E_{2x} = E_{1x} \quad \text{with} \quad z = 0$$

...Substituting these relationships in (2.35) gives

$$H_{1y} = -\frac{\sqrt{\varepsilon_c'}}{120\pi} E_{1x} \text{ A/m} \quad \text{with} \quad z = 0 \tag{2.36}$$

...which is a mathematical expression for Leontovich's approximate ...boundary conditions.

...It transpires from Leontovich's boundary conditions, the relation...ships between the horizontal components of the electric and magne...tic fields at the boundary in one medium are determined by the ...parameters of the other. Besides, the radio waves originating in ...an imperfectly conducting dielectric are plane waves propagated ...into the ground in a direction normal to the interface and experien...cing attenuation decided by the parameters of the other medium.

2.4. The Field Structure at the Point of Reception

Let there be a source located a large distance from the point of ...reception which is at the flat boundary between air and ground. ...Because of the great distance involved, the spherical wave front ...may be assumed flat within a limited space around the receiving ...aerial, and the incident wave may be treated as a plane one. In fact, ...this is an application of the same principle by which small areas ...of the earth's spherical surface are regarded as flat.

We shall use a rectangular system of coordinates in which the X axis is in the direction of propagation, and the Z axis is directed

upwards, as shown in Fig. 2.18. The electromagnetic field components in the upper medium (air) will be marked with the subscript *1*, and those in the lower (imperfectly conducting) medium, with the subscript *2*.

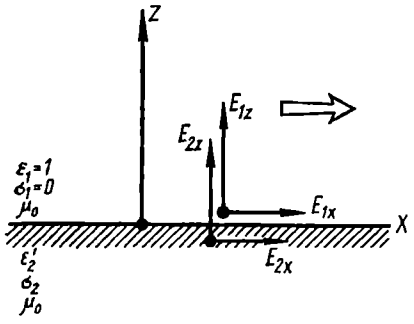


Fig. 2.18. Propagation of plane radio waves over an imperfectly conducting plane

Assuming that the vertical component of the electric field in air

$$E_{1z} = E_{1zm}e^{i\omega t} \text{ V/m} \quad (2.37)$$

is known, we seek to determine the other components in the two media.

A rigorous solution to this simple problem entails considerable mathematical difficulties. Therefore, we shall use a method based on Leontovich's approximate boundary conditions.

In cases where Leontovich's conditions are valid, a study into radio propagation along an imperfectly conducting surface may be reduced to solving Maxwell's equations for air, having imposed the above-formulated boundary conditions.

Under the circumstances, the solution of Maxwell's equations may be written as

$$E_{1z} = E_{1zm}e^{i\omega\left(t-\frac{x}{c}\right)} \text{ V/m} \quad (2.38)$$

$$H_{1y} = -\frac{E_{1zm}}{120\pi}e^{i\omega\left(t-\frac{x}{c}\right)} = -\frac{E_{1z}}{120\pi} \text{ A/m} \quad (2.39)$$

According to (2.29), the tangential components are the same on the two sides of the boundary. Therefore,

$$H_{2y} = H_{1y} \text{ A/m}, \quad \text{with} \quad z=0 \quad (2.40)$$

Substituting (2.39) in (2.36) we get

$$-\frac{E_{1z}}{120\pi} = -\frac{\sqrt{\varepsilon_c'}}{120\pi}E_{1x}, \quad \text{with} \quad z=0$$

or, in an alternate form,

$$E_{1x} = \frac{E_{1z}}{\sqrt{\varepsilon_c'}} \quad \text{with} \quad z=0 \quad (2.41)$$

Again, according to (2.29),

$$E_{2x} = E_{1x} \quad \text{with} \quad z=0 \quad (2.42)$$

substituting (1.4a) in (2.41) for the complex permittivity, we obtain

$$E_{1x} = \frac{E_{1z}}{\sqrt{\varepsilon' - i60\lambda\sigma}} \text{ V/m} \tag{2.43}$$

Equation (2.41) shows that a radio wave propagated along a flat perfectly conducting surface contains a longitudinal (horizontal) component of the electric field, that is, it is no longer a transverse wave as it would be in a homogeneous medium. The longitudinal component is ordinarily much smaller in amplitude than the vertical one and is shifted in phase relative to it.

Resolving the denominator of (2.43) into its magnitude and phase gives

$$E_{1x} = \frac{E_{1zm}}{\sqrt[4]{(\varepsilon')^2 + (60\lambda\sigma)^2}} e^{i\left(\omega t + \frac{\alpha}{2}\right)} \text{ V/m} \tag{2.43a}$$

where

$$\alpha = \text{arc tan } \frac{60\pi\sigma}{\varepsilon'} \tag{2.43b}$$

The horizontal component of the electric field decreases in amplitude with increase of ground conductivity and of wavelength. When λ tends to infinity (which occurs when a wave is propagated over perfectly conducting surface), E_{1x} tends to zero. Under ordinary conditions of ground conductivity, the horizontal component on very long, long, and medium waves is a few tenths, hundredths, or even thousandths of the vertical one.

To determine the transverse (vertical) component of the electric field in soil, we shall use the second boundary condition as expressed by (2.30). Using the adopted notation, it takes the form

$$E_{1z} = E_{2z}\varepsilon_c' \text{ (volts per metre), with } z = 0 \tag{2.44}$$

Substituting (1.4a) for ε_c' gives

$$E_{2z} = \frac{E_{1z}}{\varepsilon' - i60\lambda\sigma} \text{ (volts per metre)} \tag{2.45}$$

or in the more familiar notation,

$$E_{2z} = \frac{E_{1zm}}{\sqrt{(\varepsilon')^2 + (60\lambda\sigma)^2}} e^{i(\omega t + \alpha)} \text{ (volts per metre)} \tag{2.45a}$$

where the angle α is as given by (2.43b).

Since, according to the validity of Leontovich's conditions, the square root in (2.43a) and (2.45a) is much greater than unity, the horizontal component in the second medium is much greater than the vertical one in the wavelength bands mentioned.

For convenience, we shall write the design equations once more

$$E_{1z} = E_{1zm} e^{i\omega t} \text{ V/m} \tag{2.37}$$

$$E_{1x} = E_{2x} = \frac{E_{1zm}}{\sqrt[4]{(\varepsilon')^2 + (60\lambda\sigma)^2}} e^{i\left(\omega t + \frac{\alpha}{2}\right)} \text{ V/m} \tag{2.43a}$$

$$E_{2z} = \frac{E_{1zm}}{\sqrt{(\varepsilon')^2 + (60\lambda\sigma)^2}} e^{i(\omega t + \alpha)} \text{ V/m} \tag{2.45a}$$

$$H_{1y} = H_{2y} = -\frac{E_{1zm}}{120\pi} e^{i\omega t} \text{ A/m} \tag{2.30}$$

where

$$\alpha = \text{arc tan } \frac{60\lambda\sigma}{\varepsilon'} \tag{2.43b}$$

Since the longitudinal component of the electric field due to a radio wave propagated in air is shifted in phase relative to the vertical component of that field, the total field is elliptically polarized in a vertical plane, and the wave ceases to be the plane one. The resultant vector describes an ellipse (Fig. 2.19).

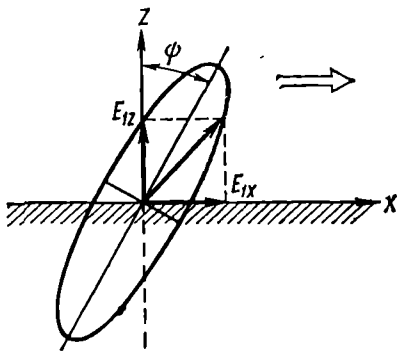


Fig. 2.19. Polarization ellipse for a radio wave propagated over an imperfectly conducting surface

Equations (2.43a) and (2.43b) show that when displacement currents prevail in the surface (which takes place when $\varepsilon' \gg 60\lambda\sigma$), the phase shift tends to zero. Conversely, when convection currents prevail, the phase shift tends to $\alpha/2 \to \pi/4$.

In the most commonly encountered types of surface, the ellipse of polarization has a very large eccentricity so that the electric field may be regarded as being plane-polarized (in the direction of propagation) and that the total field vector points in the direction of the major axis of the ellipse. Under the circumstances, the wave front may be said to be tilted. The angle of tilt of the total electric field vector relative to the normal to the interface is given by

$$\tan \psi = \frac{E_{1xm}}{E_{1zm}} = \frac{1}{\sqrt[4]{(\varepsilon')^2 + (60\lambda\sigma)^2}} \tag{2.46}$$

Example 2.5. A radio wave of $f = 1$ MHz is propagated over a moist soil such that $\varepsilon' = 10$ and $\sigma = 0.01$ S/m. The value of E_{1zm}, the vertical component of the electric field, is specified in advance. Find the other components for air and ground.

Solution: Using (2.43a) and (2.45a), we get

$$E_{1xm} = E_{2xm} \cong \frac{E_{1zm}}{13.4} = 7.45 \times 10^{-2} E_{1zm}$$

$$E_{2zm} \cong \frac{E_{1zm}}{180} = 5.5 \times 10^{-3} E_{1zm}$$

The phase difference ($\alpha/2$) between the vertical and horizontal components of the field in air is given by (2.43b)

$$\tan \alpha = 18$$

$$\alpha = 86°50'$$

$$\alpha/2 = 43°25'$$

The resultant field is elliptically polarized in the plane of propagation.

To determine the parameters of the ellipse of polarization, write the instantaneous values of the vertical and horizontal components as

$$E_{1x} = 7.45 \times 10^{-2} E_{1zm} \cos \omega t = A \cos \omega t$$

$$E_{1z} = E_{1zm} \cos (\omega t - 43°25') = C \cos \omega t + D \sin \omega t$$

where

$$A = 7.45 \times 10^{-2} E_{1zm}$$

$$C = 0.726 E_{1zm}$$

$$D = 0.687 E_{1zm}$$

The angle that the major axis of the ellipse of polarization makes with the x axis is such that

$$\tan 2\varphi = -0.107$$

whence

$$\varphi = 86°55'$$

and the angle of tilt is

$$\psi = 3°5'$$

The lengths of the major and minor axes of the ellipse of polarization are

$$M = 1.009 E_{1zm}$$

$$N = 5.12 \times 10^{-2} E_{1zm}$$

From a comparison of the axes it is seen that the ellipse is very extended.

The angle of tilt of the wave front, as found by (2.46), is

$$\psi = 4°45'$$

that is, it is very close to the angle of tilt of the major axis of the ellipse.

The tilt of the wave front is of important practical significance. For one thing, it explains why radio transmissions can be received by ground aerials (that is, balanced aerials in the form of two wires laid on the ground or strung some distance above it). Since the emf is induced in the aerial wire solely by the field component parallel to the wire, from Fig. 2.20 it follows that reception by a ground aerial is due solely to the longitudinal component of the field. If the ground were a perfect conductor, reception by a ground aerial would not be possible. On the other hand, as the ground conductivity or the wavelength decreases, the emf induced in a ground aerial increases.

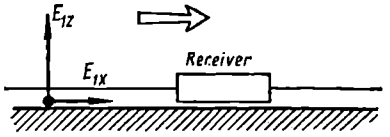


Fig. 2.20. Reception of signals by a ground aerial

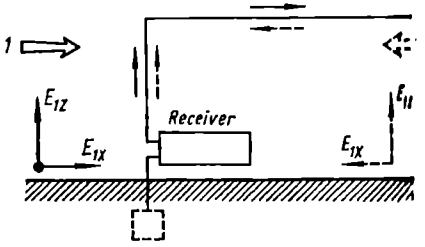


Fig. 2.21. Reception of signals by inverted-L aerial

The tilt of the wave front also explains the directional properties of inverted-L aerials widely used on long and medium waves. When used for reception, an inverted-L aerial is more effective when waves arrive along the flat-top at the side of the vertical downlead. Referring to Fig. 2.21 and assuming for simplicity that the phase difference between the vertical and horizontal components of the electric field due to the wave is negligible, the waves arriving in direction *1* induce emfs in the flat-top and downlead, which are in the same direction and add together (the full-line arrows). Conversely, when waves come from the free end of the flat-top *2*, the emfs induced in the flat-top and downlead are in anti-phase and partly cancel out (the broken-line arrows).

If an inverted-L aerial were strung over a perfectly conducting surface, it would display no directional properties. That is, inverted-L aerials owe their directivity wholly to the imperfectly conducting properties of ground. Therefore, an inverted-L aerial is more directional as the ground conductivity decreases and the wavelength received is shortened.

On the strength of the reciprocity theorem, everything said about ground and inverted-L receiving aerials fully applies to transmitting aerials.

The theory set forth above provides an insight into the reception of radio waves under the earth's surface. To begin with, let us con-

... the reception of radio waves immediately under the interface ... compare it with that immediately above the interface.

... Equations (2.43a) and (2.45a) it follows that while the verti... component of the electric field is considerably ($\sqrt[4]{(\varepsilon')^2 + (60\lambda\sigma)^2}$ ... greater than the horizontal one in air, the situation is exactly ... and under the surface. That is, for reception in air it is more ... advantageous to use vertical aerials, while reception under the ... should use horizontal aerials oriented in the direction of ... wanted station. Things are different with loop receiving aerials ... which respond to the magnetic field of the wave and in which varia... in the magnetic field may be neglected if the aerial height is ... sufficiently small. Owing to the equivalence of the magnetic field ... air and ground, a loop in air will receive the signal as effectively ... as under the surface.

... Leontovich's boundary conditions, the field at a depth h ... equal to the product of the field immediately under the interface ... the attenuation function found on the assumption that the ... is propagated in a homogeneous medium having the soil ... characteristics.

The vertical and horizontal components of the electric field at ... depth h under the interface can be found from the known vertical ... component of the electric field in air immediately above the inter... as follows

$$E_{2x} = \frac{E_{1zm} e^{-\delta h}}{\sqrt[4]{(\varepsilon')^2 + (60\lambda\sigma)^2}} e^{i\left[\omega\left(t - \frac{h}{u}\right) + \frac{\alpha}{2}\right]} \text{ V/m} \qquad (2.47)$$

$$E_{2z} = \frac{E_{1zm} e^{-\delta h}}{\sqrt{(\varepsilon')^2 + (60\lambda\sigma)^2}} e^{i\left[\omega\left(t - \frac{h}{u}\right) + \alpha\right]} \text{ V/m} \qquad (2.48)$$

... the absorption coefficient δ and the velocity of propagation ... under the surface are given by (1.27) and (1.28), and the phase ... shift α is given by (2.43b).

The magnetic field under the surface can be calculated in a simi... fashion.

The relationships established for the ground-wave field structure ... apply to distances where the effect of the earth's curvature has ... to be taken into account, because the diffracted field at the receiver ... may likewise be treated as a plane one.

Example 2.6. Determine the field at a depth of 5 m for a radio wave ... of 1 MHz, propagated over moist soil (for the soil characteristics ... of Example 2.5).

Solution: Since $60\lambda\sigma = 180$ is many times $\varepsilon' = 10$, the absorption ... factor δ should be found by Eq. (1.40)

$$\delta = 2\pi \sqrt{\frac{30 \times 0.01}{300}} = 0.198 \text{ m}^{-1}$$

Substituting the value of δ in Eq. (2.47) gives the amplitude o the horizontal (longitudinal) electric-field component of the wav

$$E_{2xm} = \frac{E_{1zm(z=0)}\, e^{-0.198\times 5}}{13.4} = 2.77 \times 10^{-2} E_{1zm(z=0)}$$

The vertical component of the electric field at the same dept can be found by Eq. (2.48)

$$E_{2zm} = \frac{E_{1zm(z=0)}\, e^{-0.198\times 5}}{180} = 2.07 \times 10^{-3} E_{1zm}$$

It is seen that in reception under the earth's surface it is mo advantageous to use aerials responding to the horizontal electri field component of the wave.

2.5. Radio Propagation over the Flat Earth, with the Aerials Located Directly at the Interface

In contrast to § 2.2, this section deals with radio propagation ov the flat earth with both aerials, transmitting and receiving, locate directly at the interface. This case is typical of very long, lon and medium waves. Since the curvature of the earth's surfa is ignored, it is clear that the calculation method involved is onl applicable to short distances.

To begin with, we shall consider a special case of the proble namely when the earth's surface may be treated as a perfect cond ctor, that is, when convection current is many times displacem current in magnitude. The assumption is quite plausible becau with very long wavelengths the ground has a fairly high conductivit Thus qualified, the problem may be solved by a rather simple metho

Assume that at point A on the flat surface of the perfectly condu ting earth there is an aerial of gain G_1 and radiated power P_1. I accordance with § 1.4, such an aerial is equivalent to an isotrop radiator of power P_1G_1. If this aerial were located in free space, t energy of the wave would be distributed over the surface of a spho In our case, radio waves are propagated through air, and the ener of the wave will be distributed over a hemisphere. As a result, t Poynting vector will be twice as great as given by Eq. (1.18a), a the field will be $\sqrt{2}$ times the value given by Eq. (1.15), or

$$S = 1.6 \times 10^5 \frac{P_{1kW} G_1}{r_{km}^2} \text{ nW/m}^2 \qquad (2.4$$

$$E_{rms} = \frac{245 \sqrt{P_{1kW} G_1}}{r_{km}} \text{ mV/m} \qquad (2.$$

Equations (2.49) and (2.50) are often referred to as the ideal ra transmission equations. This stresses a two-fold idealisation of t

problem. Firstly, we ignore the curvature of the earth's surface; secondly, the surface is treated as a perfect conductor.

Since on very long, long, and medium wavelengths use is often made of wire aerials in which the radiating element is the down-lead, it appears reasonable to derive suitable equations. This purpose can be served by the image principle known from the field theory. If above a perfectly conducting surface there is a vertical conductor in which at a given instant of time the current is flowing as shown

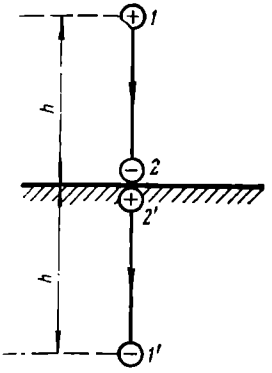


Fig. 2.22. Application of the image method to a vertical aerial above a perfectly conducting surface

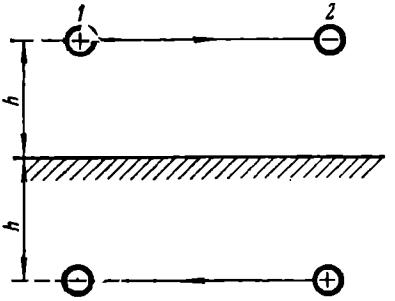


Fig. 2.23. Application of the image method to a horizontal aerial above a perfectly conducting surface

by the arrow in Fig. 2.22, this current may be visualized as being due to two charges, *1* and *2*. According to the image principle, the effect of a perfectly conducting surface on the field due to a point charge can be allowed for by replacing the surface with an image charge, that is, a charge of opposite sign located the same distance under the surface. In Fig. 2.22, the image charges are labelled *1'* and *2'*. It is easily seen that in the image conductor the current is of the same magnitude and in the same direction as in the real conductor. Thus, the effect of a perfectly conducting surface consists in doubling the field established by conductor *1-2* in free space. Noting this, we may re-write the ideal transmission equation as follows

$$E_{rms} = \frac{120\pi h_{eq(m)} I_{rms(A)}}{\lambda_m r_{km}} \text{ mV/m} \qquad (2.51)$$

where h_{eq} is the equivalent height of the transmitting aerial and I_{rms} is the rms current at an anti-node.

If the radiating wire were arranged horizontally as shown in Fig. 2.23, then, on the strength of the image principle, we would find that the current in the image conductor would be in an opposite

direction as compared with that in the real conductor. Then the image would attenuate, and not build up, the radiation in the horizontal direction. As a result, side lobes would be formed at an angle to the horizontal, in accordance with Eq. (2.20).

In considering the effect of a perfectly conducting earth, we run into a sort of paradox which consists in the following. In terms of radiated power and according to (2.50), a radiator placed above the earth increases the field $\sqrt{2}$ times in comparison with the field in free

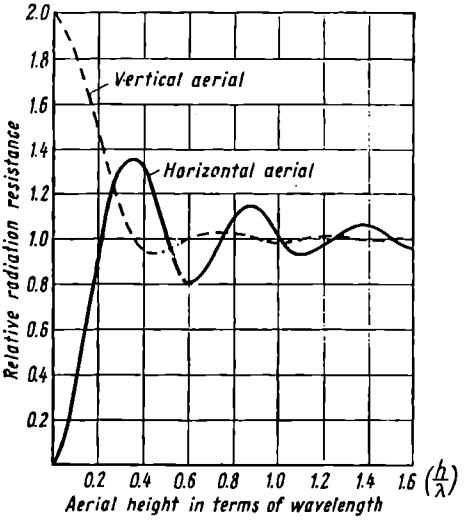


Fig. 2.24. Radiation resistance of a vertical and a horizontal elementary radiator related to that of free space as a function of radiator height above a perfectly conducting surface

space. On the other hand, in terms of current magnitude and according to (2.51), a radiator placed above the earth doubles the field as compared with the field in free space.

This seeming contradiction may be explained by the fact that different powers are required to maintain the same current in a radiator placed in free space and in a radiator placed above the earth because the radiation resistance of an aerial is a function of its height above the ground.

Fig. 2.24 shows variations in the radiation resistance of a vertical and a horizontal dipole as a function of their height above a perfectly conducting earth, relative to their radiation resistance in free space. As is seen from the plot, the radiation resistance of the vertical dipole increases as it is moved closer to the ground, doubling in the limit while that of the horizontal dipole under the same conditions decreases

tending to zero. This is an illustration of the premises shown in Figs. 2.22 and 2.23. Closer to the earth, the equivalent length of the vertical dipole tends to double, and that of the horizontal dipole tends to zero (owing to the compensating effect of the current in the image conductor).

If a dipole carrying a free-space current I_{rms} be placed vertically immediately above a perfectly conducting plane, it will take twice the previous power to maintain the same current. Replacing P_1 in Eq. (2.50) with twice its value, we find that the field at the receiver doubles in comparison with the field in free space.

For the same reason, when we express the field in terms of radiated power and compare an aerial located in free space with a vertical one placed immediately at the surface, we must remember that with the same applied power the current in the aerial will change in the ratio $1/\sqrt{2}$ of its free-space value as the aerial is moved closer to the surface.

The ideal transmission equations, (2.50) and (2.51), may be treated as the product of the free-space field and the attenuation function $F = \sqrt{2}$ for Eq. (2.50), and $F = 2$ for (2.51). In both cases, the earth builds up the field, and the attenuation function no longer lives up to its name.

As to the more practical problem, that of radio propagation over the flat imperfectly conducting earth, the first to solve it was A. Sommerfeld in 1909. Limiting himself to a surface in which the conveyance current is much greater than the displacement current, he derived exact expressions for the received field in the form of integrals. However, his solution was too unwieldy for engineering purposes. Besides, the equation contained a fundamental error rectified by a number of investigators at a later time (including V. A. Fock in 1943).

In 1923, Shuleikin [8] re-cast Sommerfeld's solution into a form suitable for engineering calculations, and in 1931, van der Pol obtained a convenient expression for the field strength almost identical with Shuleikin's.

They have been followed by other investigators who have extended Shuleikin's and van der Pol's solutions to cover surfaces with any characteristics and developed charts and nomograms to facilitate the use of the design equations.

Fig. 2.25 illustrates the statement of the problem. The transmitting aerial is set up at point A, immediately above the earth's surface. The waves radiated by the transmitting aerial are propagated over the flat homogeneous earth of permittivity ε' and of conductivity σ. We inquire about the attenuation function F for the receiving aerial located at point B, also immediately at the ground, a distance r from the radiator. Knowing the attenuation function the

field intensity can be found by the equation

$$E_{rms} = \frac{173\sqrt{P_{1\text{kW}}G_1}}{r_{\text{km}}} F \text{ mV/m} \qquad (2.$$

The attenuation function is a function of a dimensionless parameter, sometimes called the *numerical distance*. This is in turn det

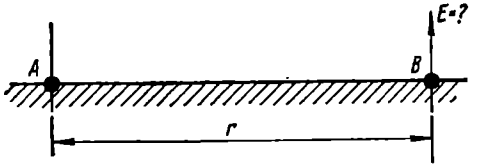


Fig. 2.25. Radio propagation over a flat earth

mined by dividing the actual distance r into the so-called *scale di tance* S which is given by

$$S = \frac{\lambda(\varepsilon_c')^2}{\pi(\varepsilon_c'-1)} \text{ m} \qquad (2.5$$

and is, in the general case, a complex quantity. In practical calculations, use is ordinarily made of its magnitude s defined as

$$s = |S| = \frac{\lambda}{\pi}\left|\frac{(\varepsilon_c')^2}{\varepsilon_c'-1}\right| \text{ m} \qquad (2.53$$

Then the numerical distance is

$$x = r/s = \frac{\pi r}{\lambda}\left|\frac{\varepsilon_c'-1}{(\varepsilon_c')^2}\right| \qquad (2.5$$

Substituting the complex permittivity in this expression gives

$$s = \frac{\lambda[(\varepsilon')^2+(60\lambda\sigma)^2]}{\pi\sqrt{(\varepsilon'-1)^2+(60\lambda\sigma)^2}} \text{ m}$$

$$x = \frac{\pi r}{\lambda}\frac{\sqrt{(\varepsilon'-1)^2+(60\lambda\sigma)^2}}{(\varepsilon')^2+(60\lambda\sigma)^2} \qquad (2.54a$$

In cases where $\varepsilon' \gg 1$, we may put $(\varepsilon'-1)^2 \cong (\varepsilon')^2$ without an impairment in accuracy. Then Eqs. (2.54a) take the form

$$s \cong \frac{\lambda}{\pi}\sqrt{(\varepsilon')^2+(60\lambda\sigma)^2} \text{ m}$$

$$x \cong \frac{\pi r}{\lambda}\frac{1}{\sqrt{(\varepsilon')^2+(60\lambda\sigma)^2}} \qquad (2.54b$$

Where the convection current in the surface is much greater tha the displacement current, that is, with $60\lambda\sigma \gg \varepsilon'$, Eqs. (2.54b

For the

$$s \cong \frac{60\lambda^2\sigma}{\pi} \text{ m}$$

$$x = \frac{\pi r}{60\lambda^2\sigma} = \frac{100\pi r_{\text{km}}}{6\lambda_{\text{m}}^2\sigma_{\text{S/m}}} \tag{2.54c}$$

For the other extreme case when the displacement current is much greater than the convection current, that is, with $60\lambda\sigma \ll \varepsilon'$ Eqs. (2.54b) reduce to

$$s \cong \frac{\lambda\varepsilon'}{\pi} \text{ m}$$

$$x \cong \frac{\pi r}{\lambda\varepsilon'} = \frac{1000\pi r_{\text{km}}}{\lambda_{\text{m}}\varepsilon'} \tag{2.54d}$$

Once the value of x is found, the value of F can be determined from the plots due to Burrows [9]. In these plots, illustrated in Fig. 2.26, the values of x are laid off as abscissa on a logarithmic scale, and the attenuation function is laid off as ordinate for two types of polarization and for different values of the parameter $Q = \varepsilon'/60\ \lambda\sigma$. As is seen, with low values of x all curves tend to $F = 1.41$. For $x > 25$, the curves also merge together.

If the plot of Fig. 2.26 is not handy, the magnitude of the attenuation function can be found approximately by the formula

$$F = 1.41\frac{2+0.3x}{2+x+0.6x^2} \tag{2.55}$$

For $x > 25$, Eq. (2.55) reduces to

$$F \cong \frac{0.707}{x} \tag{2.55a}$$

In the Soviet literature on the subject Eq. (2.52), in which the attenuation function is determined by the method just described, is known as the Shuleikin-van der Pol equation. It may be used when the receiver is a short distance from the transmitter, that is, when the curvature of the earth's surface may be neglected.

Eq. (2.54a) shows that with a perfectly conducting soil (σ is infinity), the numerical distance x reduces to zero. Therefore, the attenuation function takes on the value of 1.41 (Fig. 2.26), which corresponds to ideal radio transmission. As x increases, the attenuation function is decreased or, which is the same, the energy loss in the surface grows. From Eq. (2.54a) it also follows that the energy loss increases as the wave becomes shorter and the surface conductivity decreases.

This relationship is explained by the fact that radio waves penetrate a well conducting surface very little (in the limiting case of a perfectly conducting surface they do not penetrate at all), while

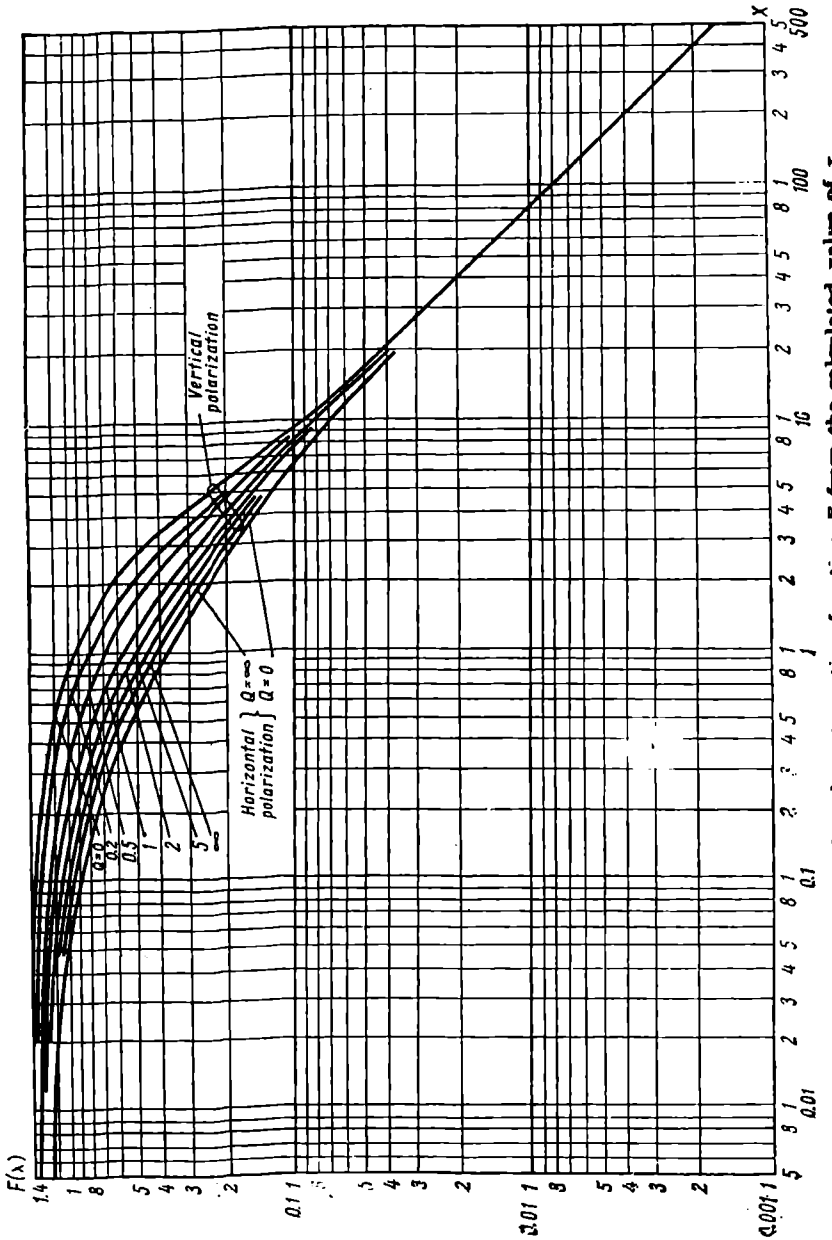


Fig. 2-26. Determining the attenuation function F from the calculated value of x

the situation reverses with poorly conducting surfaces, and the radio waves propagated above the earth contribute less and less to the total energy.

When the effective height h_e of the transmitting aerial and the aerial mode current I_{rms} are known, another form of the Shuleikin-van der Pol equation should preferably be used

$$E_{rms} = \frac{60\pi h_{e(m)} I_{rms(A)}}{r_{km}\lambda_m} F_1 \text{ mV/m} \tag{2.56}$$

where F_1 is defined as

$$F_1 = \sqrt{2}\,F \tag{2.57}$$

The limiting distances for which the Shuleikin-van der Pol equation is valid are listed in Table 2.3. They have been determined from a comparison of the calculated and measured field at the point of reception.

Table 2.3. LIMITING DISTANCES FOR SHULEIKIN-VAN DER POL EQUATION

Wave band, metres	Limiting distance, kilometres, at the mid-band
200-20,000	300-400
50-200	50-100
10-50	10

So far we have dealt with the received field due to a vertical dipole placed immediately on the imperfectly conducting surface of the earth. Naturally, such a dipole produces vertically polarized waves, and by the received field is meant the vertical component of the wave electric field.

If the source of radio waves is a horizontal dipole or any more elaborate aerial producing a horizontally polarized wave, its propagation is materially changed. For one thing, the main contribution now comes from the horizontal component of the received electric field. For another, assuming that the earth is a perfect conductor and that the horizontal dipole is placed immediately above the earth (Fig. 2.27), it may be shown on the basis of the image principle that the earth will tend to cancel the radiated field. All practical surfaces have a finite conductivity, however, and the radiated field will not be cancelled completely, that is, the horizontal component of the wave field will have a finite value at the point of reception B. In addition to this cancelling, or compensating, effect, the imperfectly conducting surface of the earth will naturally attenuate the horizontally polarized waves propagated over it.

The extent to which the horizontally polarized radio waves are attenuated is evaluated in terms of the attenuation function as found from the curves of Fig. 2.26 for horizontal polarization. The numerical distance is then given by

$$x = \frac{\pi r}{\lambda} \sqrt{(\varepsilon' - 1)^2 + (60\lambda\sigma)^2} \qquad (2.58)$$

In addition to the main horizontally polarized field whose maximum is in a direction normal to the axis of the wire, a horizontal

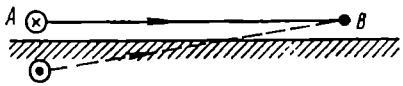


Fig. 2.27. A perfectly conducting earth compensates the radiation of a horizontal aerial. The cross and the dot mark the directions of currents in the aerial wire and its image at a given instant of time

dipole is also the source of a vertically polarized electric field whose maximum is in the direction of the wire. Fig. 2.28 shows the polar diagrams of the horizontally polarized field (the full line) and of the vertically polarized field (the broken line) due to a horizontal dipole.

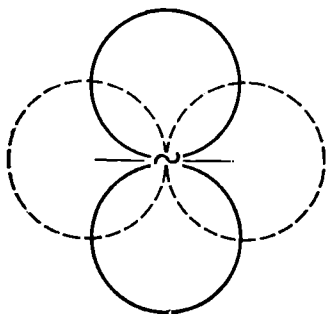


Fig. 2.28. Vertically and horizontally polarized waves due to a horizontal dipole

It should be added that a vertically polarized field appears solely due to the finite conductivity of the earth and tends to zero as the conductivity increases. The physical basis of this induced field component is the interaction between the vertical and horizontal components of the wave field. Much as a vertically polarized wave produces the horizontal field component due to the ground-induced field, so the horizontal component produces the vertical one.

The absolute value of the vertical component of the wave field can be determined from the reciprocity theorem, using the method proposed by A. N. Shchukin [10].

According to Eq. (2.55), the vertical component of the electric field at a point r distant from a vertical dipole of equivalent length l_{eq} and with an anti-node current I_{rms} is given by

$$E_{x(rms)} = \frac{60\pi l_{eq(m)} I_{rms(A)}}{\lambda_m r_{km}} F_1 \text{ mV/m} \qquad (2.59)$$

According to Eq. (2.43a), the longitudinal horizontal component of the electric field is given by

$$E_{x(rms)} = \frac{60\pi l_{eq(m)} I_{rms(A)} F_1}{\lambda_m r_{km} \sqrt[4]{(\varepsilon')^2 + (60\lambda\sigma)^2}} \text{ mV/m} \qquad (2.60)$$

On the strength of the reciprocity theorem, we may state that the same equation will describe the vertical component of the electric field, $E_{z(rms)}$, due to a horizontal dipole of effective length l_{eq} with an anti-node current I_{rms}, and at the same distance r from the dipole.

With practical surfaces, the vertically polarized field exceeds the horizontally polarized one already at small distances from a horizontal dipole placed immediately at the surface of the earth.

Example 2.7. Find the field at 250 kilometres from a transmitter, with waves propagated over a moist soil under the following conditions: $P_1 = 30$ kW, $\lambda = 1200$ m, $G_1 = 1.5$.

Solution: From Table 2.1, the surface characteristics are as follows: $\varepsilon' = 10$ and $\sigma = 0.01$ S/m, whence

$$60\lambda\sigma = 720 \gg \varepsilon' \text{ and } Q = \frac{\varepsilon'}{60\lambda\sigma} = 0.014$$

By Eq. (2.54c), the numerical distance is

$$x = \frac{\pi \times 250 \times 100}{\sqrt{6 \times 1200^2 \times 0.01}} = 0.91$$

From the plot of Fig. 2.25, the attenuation function is

$$F = 0.95$$

If we used the approximate equation (2.55), the attenuation function would be

$$F = \frac{1.41 \times 2.27}{3.41} = 0.936$$

By Eq. (2.52), the received field is

$$E_{rms} = \frac{173\sqrt{30 \times 1.5}}{250} \times 0.95 = 44 \text{ mV/m}$$

Example 2.8. Determine the current in the base of an earthed vertical aerial of $h_e = 22$ metres, with which the field at $r = 90$ km from the transmitter will be $E_{rms} = 40$ mV/m. The aerial is excited at $\lambda = 150$ m. The waves are propagated over a dry soil.

Solution: To begin with, determine the attenuation function. The soil characteristics can be taken from Table 2.1: $\varepsilon' = 4$, $\sigma = 0.001$ S/m, whence $60\lambda\sigma = 9$, and $Q = 0.444$.

Under these conditions, the numerical distance ought to be determined by Eq. (2.54a), and it will be

$$x = 184$$

Since x is large, the attenuation function may be found by the approximate equation (2.55a)

$$F \cong \frac{1}{\sqrt{2} \times 184} = 0.0040$$

and

$$F_1 = 0.0056$$

Substituting the numerical values found in Eq. (2.56) gives

$$0.0040 = \frac{60\pi \times I_{rms} \times 22 \times 0.0056}{150 \times 90}$$

whence

$$I_{rms} = 23.2 \text{ A}$$

Example 2.9. Determine the field due to a horizontal balanced half-wave dipole excited at $\lambda = 70$ m, at a distance of 30 km. The dipole is placed immediately at the surface of a moist soil. The anti-node current of the dipole is 15 A.

Solution: Because of the cancelling effect of the image radiator in the earth, the main horizontally polarized field with a maximum normal to the radiator is so small as to be negligible in comparison with the vertically polarized field.

The vertically polarized field with a maximum in the direction of the wire can be found by Shchukin's method, using Eq. (2.60). Since $60\lambda\sigma = 42$, and $\varepsilon' = 10$, it follows that $Q = \varepsilon'/60\lambda\sigma = 0.24$.

By the approximate equation (2.54a),

$$x = \frac{\pi 30 \times 100}{6 \times 49 \times 100 \times 0.01} = 32.2$$

From the plot of Fig. 2.25, the attenuation function is

$$F = 0.21 \text{ and } F_1 = 0.030$$

Substituting the numerical value of the attenuation function in Eq. (2.60) and noting that the effective length of the half-wave dipole is

$$l_{eq} = \frac{2}{\pi} \frac{\lambda}{2} = \lambda/\pi = 22.3 \text{ m}$$

the vertical component of the received field will be

$$E_{rms} = \frac{60\pi \times 22.3 \times 15 \times 0.030}{70 \times 30 \times 6.57} = 0.137 \text{ mV/m}$$

2.6. Radio Propagation over an Inhomogeneous Surface. Coastal Refraction

In practice, radio propagation over a homogeneous surface is met with seldom. More often, on their way from the transmitter to the receiver radio waves travel over surfaces widely differing

in electrical properties. Leaving out cases where surface properties change gradually, we shall consider radio propagation over a succession of surfaces (so-called *mixed-path propagation*) which remain uniform within a particular section of the path but differ from one another so that the boundaries between them are sharply defined.

To begin with, we shall consider a two-section propagation path. The solution may then be extended to cover the more general case of mixed paths with several different sections.

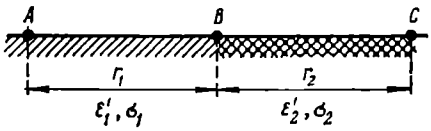


Fig. 2.29. Mixed-path propagation of radio waves (a two-section path)

The problem may be stated as follows. At point A (Fig. 2.29), there is a transmitting aerial placed at the earth surface. The radiated power P_1 and the gain G_1 of the transmitting aerial are deemed known. Part of the propagation path, r_1, is over one surface with ε_1' and σ_1. The remainder of the path, r_2, is over another type of surface with ε_2' and σ_2. We inquire about the attenuation function at point C, which is $(r_1 + r_2)$ distant from the transmitter. For convenience, the numerical distances x and the attenuation function F will be supplied with double subscripts. The first element in the subscript will denote the type of surface, (1) or (2), and the second will denote the distance that a wave covers over this type of surface. For example, the symbol x_{1r_1} stands for the numerical distance for the first type of surface, with the wave having covered a distance r_1, while the symbol $F_{2\,(r_1+r_2)}$ stands for the attenuation function of the second type of surface at a distance $(r_1 + r_2)$. The numerical distances and attenuation function will be found by the equations quoted in § 2.5.

At first, we shall determine the attenuation function by an approximate method proposed by Eckersley in 1930 [11]. The method is based on the assumption that the radio waves propagated over particular types of surface are attenuated in proportion to the distance over that surface but the attenuation is independent of the location of this surface along the path. As will be recalled, this assumption is valid for any two-wire line composed of sections differing in resistivity.

The Shuleikin-van der Pol equation applies to radio propagation over a homogeneous earth and cannot, therefore, be used directly in the case under consideration. Yet, it may be used to determine

the actual value of the attenuation function at point B, that is F_{1r_1}. Then we may find the attenuation function at the same point B on the assumption that the section AB has the properties of the other type of surface, that is, the value of F_{2r_1}, and finally the attenuation function at point C on the assumption that the entire path has the properties of the second type of surface, that is, the value of $F_{2\,(r_1+r_2)}$. The full line in Fig. 2.30 shows the attenuation function plotted against the distance over the first type of surface, and the broken line, against the distance on the assumption that

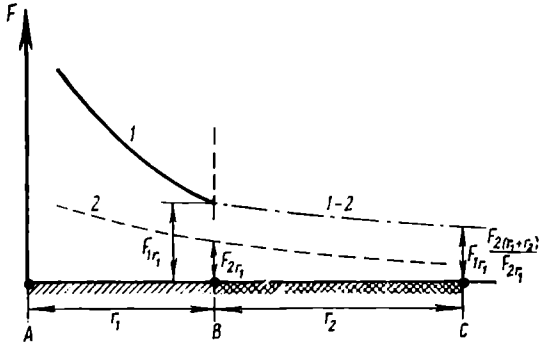


Fig. 2.30. Attenuation function versus distance for mixed-path propagation

the entire path has the properties of the other type of surface. As is seen, the attenuation function at point B is (F_{1r_1}/F_{2r_1}) times its value given by the broken line. It is natural to assume that the actual value of the attenuation function at point C is as many times its value given by the broken line. Consequently, the actual value of the attenuation function at point C is given by

$$F_C = \frac{F_{2(r_1+r_2)}F_{1r_1}}{F_{2r_1}} \tag{2.61}$$

The adjusted value of the attenuation function is shown by the dash-dot line in Fig. 2.30. Curve *1* represents radio propagation over the first type of surface; curve *2*, over the second, and curve *1—2* the adjusted attenuation function over the second section of the mixed path.

Eckersley's equation remained in use for years. In 1938, Tortladze of the Soviet Union drew this author's attention to the fact that Eckersley's method failed to satisfy the reciprocity theorem [12]. As applied to radio propagation, this theorem, widely used in the theory of linear circuits and electrodynamics, may be stated approximately as follows: If the medium in which radio waves are pro-

pagated is linear and isotropic, the attenuation function will be independent of the direction of propagation.

In our case this implies that the attenuation function at point C must have the same value as it has at point A if the transmitter is moved to point C and the receiver, to point A.

For point A the attenuation function is given by

$$F_A = \frac{F_{1(r_1+r_2)}F_{2r_2}}{F_{1r_2}} \tag{2.62}$$

which is derived from Eq. (2.61) by interchanging the subscripts 1 and 2. It has been established that the difference in value between the attenuation functions found by Eqs. (2.61) and (2.62) increases with decrease of wavelength and increase of difference between the two types of surface in conductivity. This is an indication that Eckersley's method is wrong fundamentally and that a more rigorous solution must be sought.

In 1949, Millington [13] improved upon Eckersley's method by proposing that the attenuation function be determined as the geometrical mean of F_C and F_A, that is,

$$F = \sqrt{F_C F_A} = \sqrt{\frac{F_{2(r_1+r_2)}F_{1(r_1+r_2)}F_{1r_1}F_{2r_2}}{F_{2r_1}F_{1r_2}}} \tag{2.63}$$

Naturally, with this approach the reciprocity theorem is satisfied automatically, but the method itself is not more rigorous. It should be noted, however, that Millington's method yields fairly good results for most practical problems. Some deviations from the actual field strength may occur only immediately at the interface.

From Eq. (2.63) it follows that for its practical application we must know all values of the attenuation function; in our case, it may take on six values. In cases where the numerical distances for the two sections of the path are much greater than unity, that is, when the attenuation function may be found approximately by Eq. (2.55a), Eq. (2.63) may be markedly simplified. Writing then the attenuation functions as

$$F_{1r_1} = \frac{0.707}{x_{1r_1}} = \frac{0.707s_1}{r_1}$$

$$F_{2r_2} = \frac{0.707s_2}{r_2}$$

$$F_{2r_1} = \frac{0.707s_2}{r_1}$$

$$F_{1r_2} = \frac{0.707s_1}{r_2}$$

and substituting them in Eq. (2.63), we get

$$F = \sqrt{F_C F_A} = \sqrt{\frac{F_{2(r_1+r_2)} F_{1(r_1+r_2)} 0.707 s_1 \times 0.707 s_2 \times r_1 r_2}{r_1 r_2 \times 0.707 s_2 \times 0.707 s_1}} =$$

$$= \sqrt{F_{2(r_1+r_2)} F_{1(r_1+r_2)}} \qquad (2.64)$$

Eq. (2.64) can be readily generalized to propagation over paths having three and more sections.

A rigorous study into ground-wave propagation over an inhomogeneous earth (the mixed-path problem) has been carried out by

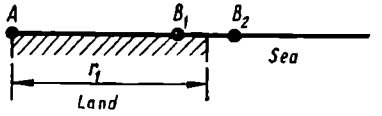


Fig. 2.31. An increase in the field strength as the receiver is moved from B_1 to B_2

Grinberg [14], Fock [15], and Feinberg [16], all of the Soviet Union. It is worth while to recapitulate, at least in brief, the last-named work.

As in the investigation of the wave structure at the point of reception and in the derivation of the Shuleikin-van der Pol equation, it is assumed that the magnitude of the complex permittivity is much greater than unity. Using Leontovich's approximate boundary conditions, Feinberg arrived at an exact expression for the attenuation function in the case of a vertically polarized wave propagated along a mixed path.

We shall quote two of Feinberg's equations [17]. One applies to radio propagation partly over land, r_1, and partly over sea, r_2, whose conductivity is assumed to be infinitely large. Assuming further that the numerical distance for the land section of the path is much greater than unity, that is, $x_{1r_1} \gg 1$, the attenuation function is

$$F = \frac{0.707}{x_{1(r_1+r_2)}} \sqrt{1 + \frac{4}{\pi} x_{1(r_1+r_2)} \frac{r_2}{r_1}} \qquad (2.65)$$

In radio propagation over a homogeneous earth, the attenuation function always decreases with increasing distance, while there may be exclusions from this rule along mixed paths. An example is offered by the case diagrammatically shown in Fig. 2.31, where the transmitter is located at A on land, a distance r_1 from the coastline, and the receiver is moved from point B_1 (near the water edge, but still on land) to point B_2 (near the water edge, but in the sea). Assuming that $x_{1r_1} \gg 1$, the attenuation function for

point B_1 may be found by Eq. (2.55a), that is,

$$F_{B_1} \cong \frac{0.707}{x_{1r_1}}$$

(assuming that the distance to point B_1 is practically equal to r_1). The attenuation function at point B_2 may be found by Eq. (2.65), on neglecting r_2 in comparison with r_1 in $x_{1\,(r_1+r_2)}$. Then,

$$F_{B_2} \cong \frac{0.707}{x_{1r_1}} \sqrt{1 + \frac{4}{\pi} x_{1r_1} \frac{r_2}{r_1}}$$

Since the radicand is undoubtedly greater than unity, the attenuation function at point B_2 and, consequently, the wave field may be greater than they are at point B_1 under certain conditions.

The field strength distribution along a mixed path with a total length of 150 kilometres and for a transmitted power of one kilowatt

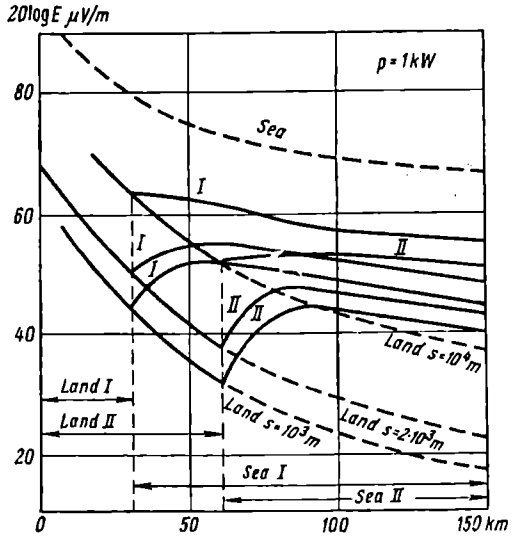


Fig. 2.32. Field strength distribution along a "land-sea" path for 1 kilowatt of radiated power and three values of $s = 60\lambda^2\sigma/\pi = 10^4$, 2×10^3 and 10^3 m; I—$r_1 = 30$ km; II—$r_1 = 60$ km

is shown in the plot of Fig. 2.32 taken from Feinberg's publication. Curves I apply to a land section 30 kilometres long, and curves II to a land section 60 kilometres long. The curves are plotted for three values of the scale distance s, as found by Eq. (2.53a) which reduces to

$$s = \frac{60\lambda^2\sigma}{\pi}$$

when $60\lambda\sigma \gg \varepsilon'$. The three values used in the plot of Fig. 2.32 are 10^4, 2×10^3, and 10^3 metres. The broken lines represent the field distribution for paths wholly over sea or land.

As follows from Fig. 2.32, as the wave crosses the coast line, the field strength increases, in full agreement with the foregoing (Fig. 2.31), in all cases except curve *I* for $s = 10^4$ metres.

The other of Feinberg's equations gives the attenuation function in the general case when both sections of the path have a finite conductivity. If both sections have large numerical distances, that is, if $x_{1r_1} \gg 1$ and $x_{2r_2} \gg 1$, the general equation takes an especially simple form

$$F \cong \frac{0.707\sqrt{s_1 s_2}}{r_1 + r_2} \tag{2.66}$$

where $s_1 = 60\lambda^2\sigma_1/\pi$ is the scale distance for one path section, and $s_2 = 60\lambda^2\sigma_2/\pi$ is the scale distance for the other.

Placing $(r_1 + r_2)$ under the radical sign, re-writing Eq. (2.66) as

$$F \cong \frac{0.707}{\sqrt{\frac{r_1+r_2}{s_1}\frac{r_1+r_2}{s_2}}} \tag{2.66a}$$

and recalling the relationship between the scale distance s and the numerical distance x (Eqs. 2.54), we may re-write Eq. (2.66a) as

$$F \cong \frac{0.707}{\sqrt{x_{1(r_1+r_2)}x_{2(r_1+r_2)}}} \tag{2.66b}$$

As follows from Eq. (2.55a) when $x \gg 1$, $F \cong 0.707/x$, so that Eq. (2.66b) may be re-written as

$$F \cong \sqrt{F_{1(r_1+r_2)}F_{2(r_1+r_2)}} \tag{2.66c}$$

which checks with Eq. (2.64) derived earlier.

Eq. (2.66c) shows that in our case the attenuation function is the geometrical mean of the attenuation functions found for homogeneous paths of length $(r_1 + r_2)$, having the properties of the two types of surface.

Feinberg has also proposed a method for the calculation of the attenuation function for radio waves propagated over three-section paths, with two sections (adjacent to the transmitting and receiving aerials) assumed identical in properties and having the same length.

The surfaces chosen are sea (as a well conducting medium causing an insignificant attenuation to radio waves) and land (as a poorly conducting medium causing a heavy attenuation to radio waves). Writing r_1 for the length of the path over land, and r_2 for the length of the path over sea, the total path length is $r_1 + r_2$. Let us agree to call the ratio $\xi = r_1/(r_1 + r_2)$ the *land-fill factor* of the path.

Then the attenuation function may be plotted against the land-fill factor as shown in Fig. 2.33. The upper curve represents cases where the sea sections are adjacent to the transmitter and receiver, and the lower curve, when the land sections are adjacent. In the calculations it is assumed that sea water has an infinitely large conductivity.

A study of the plot will show that when the land sections are adjacent to the transmitter and receiver, even small tracts of land (with a land-fill factor of about 0.1) cause a sudden and marked decrease in the attenuation function. Conversely, when the sea sections are adjacent to the transmitter and receiver and the land may be pictured as a small island somewhere midway along the path, an increase in the land section will only cause a gradual and insignificant decrease in the attenuation function. A sudden decrease in the field strength will occur only when the land sections are close to the transmitter and receiver (with a land-fill factor of over 0.9).

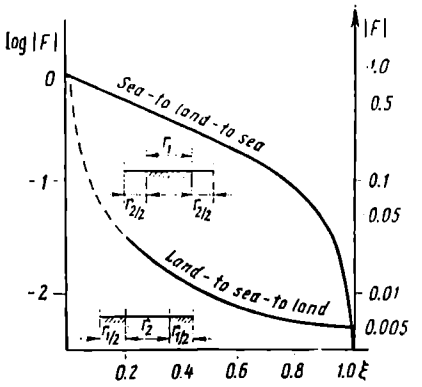


Fig. 2.33. Attenuation function of radio propagation over sea and land versus "land-fill" factor

The example just discussed clearly shows that as regards ground waves the attenuation caused by a particular segment of the path depends on its location along the path. The heaviest attenuation is caused by the segments immediately adjacent to the transmitting and receiving aerials, aptly called by Mandelstamm "take-off" and "touch-down" terminals. Furthermore, it clearly reveals the nonrigorous character of Eckersley's method which is based, as already noted, on the analogy with a two-wire line where the location of attenuating segments does not affect the resultant attenuation.

Physically, this behaviour of ground waves may be explained by the fact that near the transmitting and receiving aerials the ground waves are forced to propagate in close proximity to the earth and are subject to the absorbing effect of the ground to the full extent. On moving away from the transmitter, the waves lift themselves, as it were, from the ground and are propagated in the free atmosphere some distance above the interface.

According to Huygens' principle, each point on the wave front is the source of secondary waves which break away from the main flow of energy and reach the earth's surface. In a very rough manner,

this is shown in Fig. 2.34 where the full line shows the main flow of energy, and the broken lines represent the secondary waves moving towards the ground.

This propagation of radio waves along a trajectory of minimum absorption is not unexpected. An illustration is offered by radio communication between two submerged submarines*. Let the

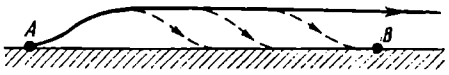


Fig. 2.34. Propagation of surface waves over areas adjacent to the transmitting and receiving aerials

distance r between the submarines be much greater than their depth of submergence h (Fig. 2.35). From § 1.5 it is known that radio waves propagated in sea water experience a strong absorption. Therefore, the waves propagated along the shortest and, it would

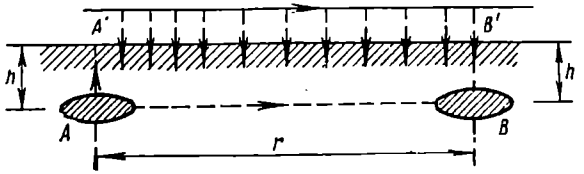


Fig. 2.35. Pertaining to the definition of a "take-off" and a "touch-down" terminals

seem, natural path AB would produce a negligible field at the point of reception. A stronger field would be due to the wave coming out along the shortest path to the water surface at point A', propagating like a ground wave along the interface, and diving at point B' to depth h to reach the submarine B. Of course, flows of energy would travel into the sea all the way along their path, as shown in Fig. 2.35. The path $AA'B'B$ resembles very closely the path with the "take-off" and "touch-down" terminals shown in Fig. 2.34.

Example 2.10. Plot the attenuation function against distance, if a wave of 227 m is propagated over a dry soil with $\varepsilon' = 4$ and $\sigma = 0.001$ S/m for the first 60 km, and then over sea water whose conductivity is assumed to be infinitely large.

Use the exact equation (2.65) and the approximate equation (2.64) in the range of distances from 30 to 100 km.

---

* The situation is purely illustrative. Actual communication would be by more efficient methods.

Solution: Since $60\lambda\sigma = 60 \times 227 \times 0.001 =$ approx. 14 is greater than the permittivity ($\varepsilon' = 4$), the scale distance may be found from the simplified expression

$$s = 60\lambda^2\sigma/\pi$$

Substituting the values of λ and σ, we get

$$s = 10^3 \text{ m}$$

Let us use the Shuleikin-van der Pol equation in order to find the attenuation function for distances of 30 and 60 kilometres. For $r = 30$ km,

$$x = \frac{r}{s} = \frac{30 \times 10^3}{10^3} = 30$$

Eq. (2.55a) gives

$$F = \frac{0.707}{x} = 0.0236$$

Similarly, for $r = 60$ km,

$$F = \text{(approx.) } 0.0118$$

Use the exact equation in order to find the attenuation function for distances of 65, 70, 100 kilometres. Substituting the respective values in Eq. (2.65) yields:

$$F = 0.0306 \text{ for } r = 65 \text{ km}$$

$$F = 0.0403 \text{ for } r = 70 \text{ km}$$

$$F = 0.072 \text{ for } r = 100 \text{ km}$$

Now use Eq. (2.54a) for the same distances. According to the statement of the problem, the conductivity of sea water is assumed to be infinitely large. Therefore, F_2 will in all cases be 1.41 very nearly. Then Eq. (2.54a) will take the form

$$F \cong \sqrt{\frac{1.41 F_{1(r_1+r_2)} F_{1r_1}}{F_{1r_2}}}$$

For $r = 65$ km

$$F = \sqrt{\frac{1.41 \times 0.707 \times 0.707 \times 5}{65 \times 60 \times 0.707}} = 0.0355$$

Similarly, for $r = 70$ km, $F = 0.0488$, and for $r = 100$ km, $F = 0.082$.

The results are presented in the plot of Fig. 2.36. As is seen, the exact and the approximate methods check fairly well even when there is a good deal of difference between the two types of surface, which fact justifies the use of the approximate method for engineering purposes.

The general expression derived by Feinberg for the attenuation function for mixed-path propagation, notably for "land-to-sea" paths, may be applied to the study of coastal refraction.

The name "coastal refraction" is applied to the bending of the path of a direct radio wave when it crosses the coast. Coastal refraction was noticed in operation of shore-based direction-finding

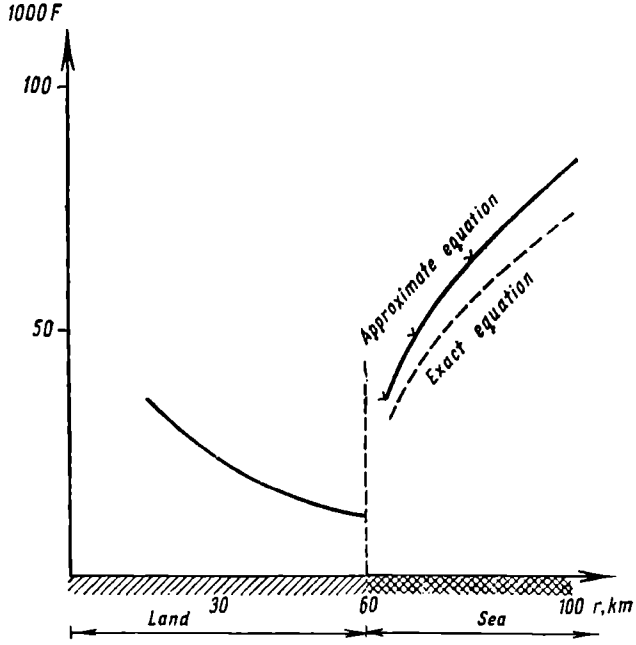


Fig. 2.36. Attenuation function for mixed-path propagation, as found by the exact and the approximate equations

stations as early as 1918 and was found to be the source of direction-finding errors. In 1919, Eckersley, basing himself on the theory advanced by Zenneck [18] to explain the propagation of surface waves, formulated a theory of coastal refraction. Solving Maxwell's equation for radio waves propagated along the interface between air and land and working from the exact boundary conditions, Zenneck derived an expression for the velocity of propagation and for attenuation functions both in the direction of propagation and in a direction normal to it (that is, for waves penetrating the ground and moving upwards).

At that time, Zenneck's theory appeared valid for large distances from the transmitter, when the received wave can be treated as

a plane one, and Eckersley naturally based himself on it in his attempts to investigate direction-finding errors. Using the classical law of refraction and substituting in it the phase velocities over land and sea, which exceeded that of light in a vacuum in Zenneck's theory, T. L. Eckersley arrived at very simple formulae for the angle of deviation from the original direction. However, the angles of deviation calculated by Eckersley's formulae failed both to agree with deviations observed experimentally and to indicate the true direction of deviation.

Several years later it was found [19] that the usual aerials could not excite Zenneck waves. To produce them, use must be made of special aerial arrays. Today, these waves, now called "surface", are widely used in many applications. The invalidity of Zenneck's theory to ground waves explains why Eckersley failed in his theory.

A consistent theory of coastal refraction and a method for the calculation of direction-finding errors have been elaborated by Grinberg, Fock, and Feinberg of the Soviet Union. Because of the complexity of the mathematics used in their work, we shall limit ourselves to the more general findings of the study.

Their approach to coastal refraction may be outlined as follows.

Assume that at point A (Fig. 2.37) there is a ship-borne transmitter. Its signals are picked up by a direction-finder placed at point B, some distance from the coast line. Without coastal refraction the radio waves would be propagated along the normal n to the equi-phase line which would be an arc of a circle with the transmitter as centre. Because of coastal refraction, however, the symmetry about point A is upset, and the equi-phase line takes up a different position shown by the broken line in Fig. 2.37, and the actual direction of propagation at the direction finder is represented by the normal m. The angle α between the two normals, n and m, is the direction-finding error caused by coastal refraction.

Among the factors causing the equiphase line to depart from the great-circle path is this. Until now, in dealing with the propagation of ground waves over inhomogeneous surfaces, we were interested solely in the absolute value of the field strength. In other words, we took into account only the magnitude of the attenuation function. It can be readily shown that the phase of the attenuation function likewise affects the phase of the radio wave, as it does in the case of radio propagation over homogeneous surfaces.

Referring to Fig. 2.37, it will be noted that path AB differs from path AB' (points B and B' are on a circle of radius r and with point A as centre) in the value of the land-fill factor. As a consequence, both the amplitude and the phase of the radio wave will be different

at points B and B'. Assume that the wave propagated in the direction AB' has the same phase at point B' and at point B''. The evaluation of the direction-finding error reduces to finding the position of the equiphase line on land, that is, in the final analysis, to determining the phase of the wave propagated along a mixed path and crossing the coast.

To find the angle α, which will be taken as positive when the angle of refraction is greater than the angle of incidence, it is convenient to introduce two systems of Cartesian coordinates on the

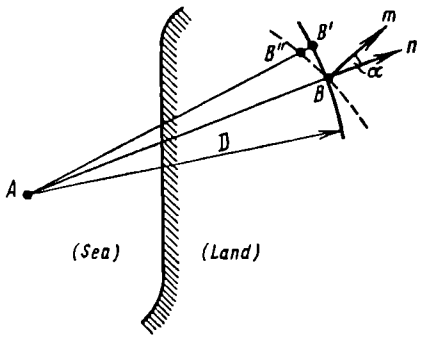


Fig. 2.37. Coastal refraction

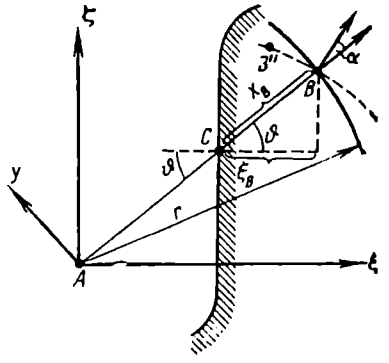


Fig. 2.38. Angular error due to coastal refraction

plane (a sea chart), as shown in Fig. 2.38. The X-axis of one is directed along the straight line connecting transmitter A and direction-finder B. The ξ-axis of the other is parallel with the normal to the coast line at the point of incidence C. The origin of coordinates is at point A in both cases.

To simplify calculations, Feinberg has compiled a nomogram (reproduced in Fig. 2.39), which is fairly simple to use. One of the three curves is selected by the ratio r/ξ_B. Then, taking the specified angle of incidence, the direction-finding error in degrees is found from one of the 18 ordinates corresponding to the specified combination of λ, ξ_B and the electric characteristics of the land.

Direction-finding errors appear on distances such that

$$x_B \ll \frac{\lambda |\varepsilon_c'|}{\pi} \text{ m}$$

On larger distances, direction-finding errors do not usually exceed several degrees.

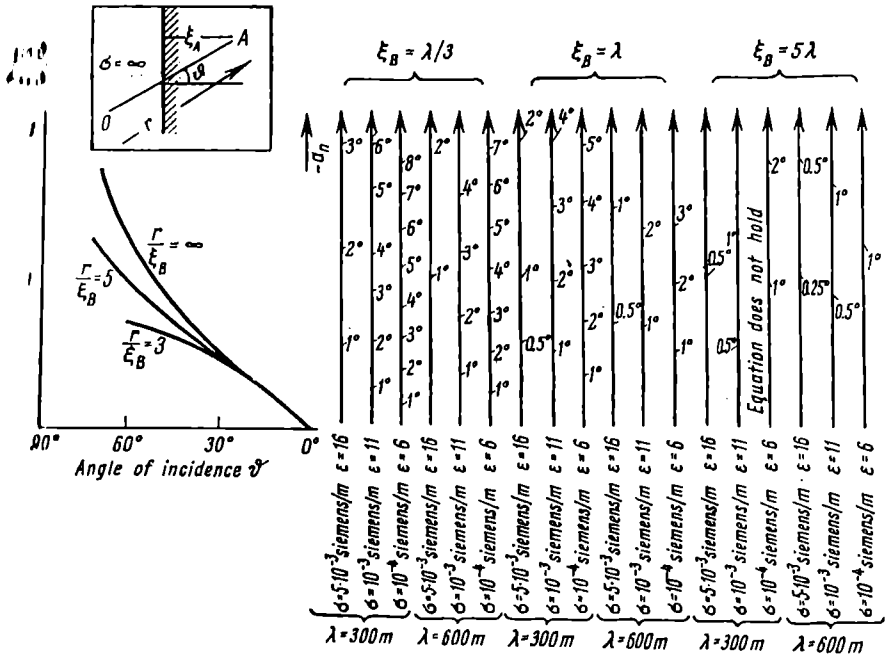


Fig. 2.39. A nomogram to find the direction-finding error due to coastal refraction

A more detailed discussion of the ground-wave field structure will be found in [20]. The velocity of propagation of ground waves is discussed in [21, 22, 23].

2.7. Spherical-earth Problem. Line-of-Sight Propagation. Effect of the Earth's Curvature

Line-of-Sight Range. The crucial question arising in a study into radio propagation over the actual (that is, spherical) surface of the earth is whether or not the transmitting and receiving aerials are within the line-of-sight range of each other.

If one of the aerials, A, is elevated, and the other, C, is on the ground (Fig. 2.40), the problem reduces to finding the distance to the visible horizon. Setting here and elsewhere the radius of the earth equal to $a = 6.37 \times 10^6$ metres from the right triangle OAC we have

$$\cos \alpha = \frac{a}{a + h_1} \cong 1 - \frac{h_1}{a} \qquad (2.67)$$

where we have left only the first two terms of the series because the height h_1 is negligible in comparison with the earth's radius.

In all practical problems, the geocentrical angle is very small, and we may therefore write

$$\cos \alpha \cong 1 - \frac{\alpha^2}{2} \tag{2.68}$$

Equating (2.67) and (2.68) gives

$$\alpha = r_{10}/a = \sqrt{\frac{2h_1}{a}}$$

whence the horizon distance is

$$r_{10} = \sqrt{2ah_1} \text{ m} \tag{2.69}$$

Substituting the numerical value of a and writing r_{10} in kilometres, we get

$$r_{10} = 3.57 \sqrt{h_{1(m)}} \text{ km} \tag{2.69a}$$

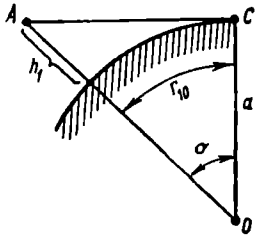


Fig. 2.40. Determining the line-of-sight range when one of the aerials is on the ground

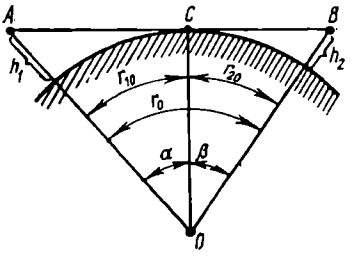


Fig. 2.41. Determining the line of-sight range with both aerials elevated

Eq. (2.69a) can be readily generalized to two elevated aerials (Fig. 2.41). Noting that the straight line AB touches the ground at point C, we write

$$r_0 = r_{10} + r_{20} = \sqrt{2a}\,(\sqrt{h_1} + \sqrt{h_2}) \text{ (metres)} \tag{2.70}$$

or

$$r_0 = 3.57\,(\sqrt{h_{1(m)}} + \sqrt{h_{2(m)}} \text{ (kilometres)} \tag{2.70a}$$

Allowance for the Earth's Curvature in the Use of the Interference Equations. We shall limit ourselves to cases where the path length r (the distance between transmitter and receiver) is shorter than the horizon distance.

The propagation of ground waves with both the transmitting and receiving aerials elevated is shown in Fig. 2.42. As with waves propagated over a flat earth, the total field at the receiver should be regarded as the sum of the direct ray AB and the reflected ray ACB.

The curvature of the earth's surface has a two-fold effect on radio propagation within the line of sight. For one thing, with the fixed heights of the transmitting and receiving aerials the magnitude of the path length difference between the direct and reflected ray will be other than that found by Eq. (2.14) for the flat earth.

For another, after reflection at the convex surface, the ray paths will diverge more than before reflection, and this will reduce the power received via the reflected ray.

The propagation of a direct and a reflected ray over the spherical earth is shown diagrammatically in Fig. 2.42. If we draw a plane MN tangent to the earth at the point of reflection and measure the aerial heights from this plane instead of from the earth's surface, we shall get so-called reduced aerial heights, h_1' and h_2'. Substituting them for the actual heights h_1 and h_2 in Vvedensky's equation, we shall get a correct value of the path length difference and, consequently, a correct value of the received field, because the angle of elevation of the ray above the convex surface of the earth is the same as the angle of elevation above the tangent plane.

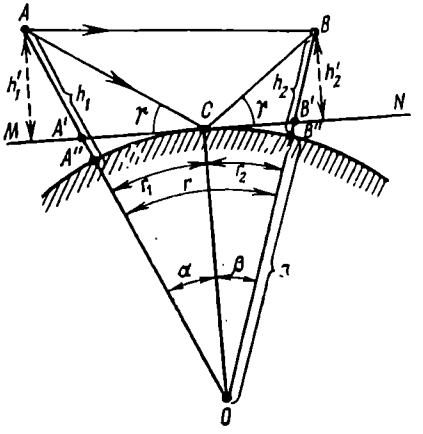


Fig. 2.42. Finding the path-length difference for radio propagation over a spherical earth

In other words, we may extend all methods for field calculation at the receiver discussed in the previous section to the spherical earth, provided the actual aerial heights are replaced with their reduced values.

For example, when using the interference equations, the effect of the earth's curvature can be allowed for by finding the reduced heights h_1' and h_2' from the known r, h_1, and h_2.

Before deriving expressions for h_1' and h_2', it should be noted that Fig. 2.42 does not present these quantities in a true proportion, because the heights are shown on an exaggerated scale for clarity. Actually, there is practically no divergence between h_1 and h_1', and we may put that

$$\left.\begin{aligned} h_1' &= h_1 - \Delta h_1 \\ h_2' &= h_2 - \Delta h_2 \end{aligned}\right\} \qquad (2.71)$$

where Δh_1 and Δh_2 represent the straight-line segments $A''A'$ and $B''B'$.

Noting that r_1 represents the line-of-sight range, or the horizon distance, at height h_1, from Eq. (2.69) we find

$$\Delta h_1 = \frac{r_1^2}{2a} \quad \text{and} \quad \Delta h_2 = \frac{r_2^2}{2a}$$

Substituting in Eq. (2.71) gives

$$h_1' = h_1 - \frac{r_1^2}{2a} \quad \text{and} \quad h_2' = h_2 - \frac{r_2^2}{2a} \tag{2.72}$$

Setting $a = 6.37 \times 10^6$ metres and expressing r in kilometres, we get

$$\begin{aligned} h_1' &= h_{1(m)} - \frac{r_{1(\text{km})}^2}{12.8} \text{ m} \\ h_2' &= h_{2(m)} - \frac{r_{2(\text{km})}^2}{12.8} \text{ m} \end{aligned} \tag{2.72a}$$

From Eqs. (2.72a) it is seen that in order to find the reduced aerial heights it is important to know the position of the reflection point C, or, which is the same, the distances r_1 and r_2.

Referring to Fig. 2.42, the angle of incidence at point C is equal to the angle of reflection, so that from the triangles OAC and OCB we may write

$$(a + h_1)\cos(\gamma + \alpha) = a\cos\gamma$$
$$(a + h_2)\cos(\gamma + \beta) = a\cos\gamma$$

Finding $\tan\gamma$ (where γ denotes the grazing angles ACM and BCN) from each equation and equating, we get

$$\tan\gamma = \frac{\cos\alpha - \dfrac{a}{a+h_1}}{\sin\alpha} = \frac{\cos\beta - \dfrac{a}{a+h_2}}{\sin\beta} \tag{2.73}$$

Eq. (2.73) is a quite rigorous one, because no assumptions were made in its derivation. Although it cannot be solved analytically, a graphical solution can be found.

Eq. (2.73) may be simplified without any detriment to the required accuracy of calculations. Noting that $a \gg h_1$ and $a \gg h_2$, we may set

$$\frac{a}{a+h_1} \cong 1 - \frac{h_1}{a} \quad \text{and} \quad \frac{a}{a+h_2} \cong 1 - \frac{h_2}{a}$$

Since the geocentric angles are small, the sines may be replaced with their arguments, and we may write on the same grounds

$$\cos\alpha \cong 1 - \frac{\alpha^2}{2}$$

and

$$\cos\beta \cong 1-\frac{\beta^2}{2}$$

so that Eq. (2.73) may be re-written as

$$\tan\gamma=\frac{h_1/a-\alpha^2/2}{\alpha}=\frac{h_2/a-\beta^2/2}{\beta}$$

Expressing the geocentric angles in terms of the respective distances r_1 and r_2 we get

$$\tan\gamma=\frac{h_1-r_1^2/2a}{r_1}=\frac{h_2-r_2^2/2a}{r_2} \tag{2.74}$$

such that $r_1+r_2=r$.

Setting $h_1>h_2$ and solving Eq. (2.74) for r_1, we get

$$r_1=\frac{r_2}{2}+2\sqrt{\frac{r^2}{12}+\frac{a}{3}(h_1+h_2)}\times$$

$$\times\cos\left\{60^\circ+\frac{1}{3}\arccos\frac{ar(h_1-h_2)}{4\left[\frac{r^2}{12}+\frac{a}{3}(h_1-h_2)\right]^{3/2}}\right\} \tag{2.75}$$

When the propagation path (distance between transmitter and receiver), r is much shorter than the horizon distance r_0, the reflection point is located by the equations for a flat earth which have the form

$$r_1=\frac{h_1}{h_1+h_2}r$$

and (2.76)

$$r_2=\frac{h_2}{h_1+h_2}r$$

In practical calculations based on the interference equations that contain the path length difference Δr, the reflection coefficient magnitude R, and the reflection coefficient phase θ [such as Eq. (2.9)], it is important to remember that the geometrical heights h_1 and h_2 should be likewise replaced with their reduced values in the expressions for the path length difference [(2.14)] and for the grazing angle [(2.11)] which is used as the argument in finding R and θ from the plots.

After such replacements Eqs. (2.11) and (2.14) take, respectively, the form

$$\gamma\cong\frac{h_1'+h_2'}{r} \tag{2.77}$$

$$\Delta r\cong\frac{2h_1'h_2'}{r}\,\mathrm{m} \tag{2.78}$$

This takes care of the effect that the spherical earth has on the path length difference and the grazing angle. As to the increased divergence of the reflected rays, it is insignificant and may be neglect-

ed in engineering calculations. The errors due to an inaccurate knowledge of surface characteristics and surface irregularities more than absorb the effect of the ray divergence.

It is not difficult to note that the complete interference equation (2.9a) and its simplified forms (2.16) and (2.22) contain the product of the actual aerial heights, h_1h_2. Therefore, in practical solution of the spherical-earth problem we need not determine the reduced height for each aerial separately. It will suffice to find the product of the reduced heights, $h_1'h_2'$. This can be done by use of the plot shown in Fig. 2.43 due to [24]. It gives the correction factor m in the equation

$$h_1'h_2' = mh_1h_2\,\mathrm{m}^2 \tag{2.79}$$

In the plot of Fig. 2.43, the ratio h_2/h_1 (where h_1 is the larger height, irrespective of whether this is a transmitting or a receiving aerial) is plotted as abscissa. According to the reciprocity theorem, there must be identical conditions for transmission in either way. The independent variable is

$$p = \frac{r}{\sqrt{2ah_1}} \tag{2.80}$$

The values of the correction factor m are then taken along the axis of ordinates.

In a similar way and using the plot of Fig. 2.44 we may find the correction factor n for the equation of the grazing angle

$$\gamma \cong n\frac{h_1+h_2}{r} \tag{2.81}$$

Example 2.11. Using the plot of Fig. 2.43, determine the path length difference for the spherical earth under the conditions defined in Example 2.3.

Solution: The ratio of the aerial heights is

$$\frac{h_2}{h_1} = \frac{10}{25} = 0.4$$

The independent variable

$$p = \frac{10^4}{\sqrt{2\times 6.37\times 10^6\times 25}} = 0.57$$

Referring to the plot of Fig. 2.43, $m = 0.79$.

Substituting the values of h_1, h_2, and m in Eq. (2.79) gives

$$h_1'h_2' = 0.79 \times 25 \times 10 = 197.5\ \mathrm{m}^2$$

By Eq. (2.78),

$$\Delta r = \frac{2\times 197.5}{10^4} = 0.039\ \mathrm{m}$$

The Range of Validity of the Interference Equations. Referring to Fig. 2.42, it is not difficult to note that as the path length r

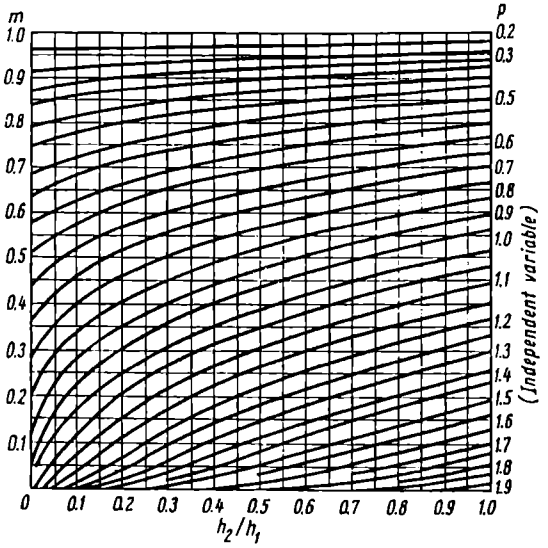


Fig. 2.43. The correction factor $m = h_1' h_2'/h_1 h_2$ as a function of $h_2 h_1$ and $p = r/\sqrt{2ah_1}$

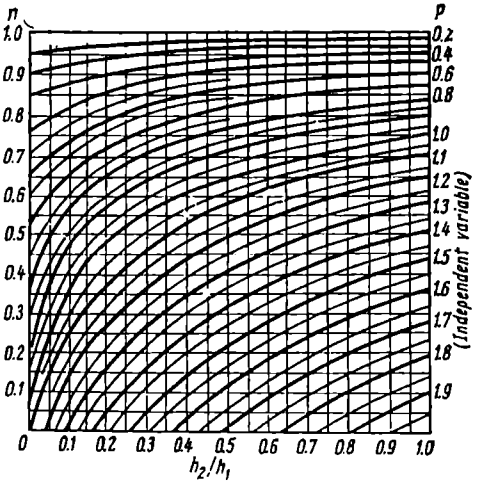


Fig. 2.44. The correction factor n as a function of h_2/h_1 and $p = r/\sqrt{2ah_1}$

increases and approaches the horizon distance, the reduced aerial heights h_1' and h_2' gradually decrease, tending to zero in the limit (with $r \to r_0$). Under the circumstances, all forms of the interference equations reduce to Vvedensky's equation (2.22) which then takes the form

$$F = \frac{4\pi h_1' h_2'}{\lambda r} = \frac{4\pi m h_1 h_2}{\lambda r} \tag{2.82}$$

It is clear then that as we approach the horizon distance, the attenuation function and, as a consequence, the field strength E_{rms} reduces to zero.

The same result is obtained if we use the plots of Fig. 2.43 and Fig. 2.44, which are so constructed that as r approaches r_0 the factor m tends to zero. As a proof, noting the expression for the independent variable p, Eq. (2.80), and using Eq. (2.70), we obtain

$$\frac{r_0}{\sqrt{2ah_1}} = 1 + \sqrt{\frac{h_2}{h_1}}$$

or

$$p = 1 + \sqrt{\frac{h_2}{h_1}}$$

When $h_2 = h_1$, $p = 2$, which corresponds, as follows from Fig. 2.43, to m tending to zero. In the other limiting case when $h_2 = 0$, $p = 1$, which also corresponds to m tending to zero.

The statement that the attenuation function tends to zero on ranges comparable with the horizon distance is not true. Experience shows that the ground-wave field does exist at distances much greater than the distance to the visible horizon. The invalidity of Eq. (2.82) for distances $r \to r_0$ implies that the underlying concept, namely that the field at the reflection point is due to the combination of the direct and reflected rays, looses its meaning under the circumstances.

The fact is that within the line-of-sight range the direct and reflected waves merge together, and the total field can only be determined by the more rigorous methods of electrodynamics instead of the crude techniques of geometric optics. Among other things, this can be done by the diffraction equations discussed in the following section. In approximate calculations, the interference equations may be used on ranges up to $r \leqslant 0.7r_0$ to $0.8r_0$.

2.8. Radio Propagation over a Smooth Homogeneous Spherical Earth

In determining the effect of the earth's curvature on radio propagation with elevated transmitting and receiving aerials, it is customary to divide the propagation path into three sections: an illumi

nated zone, a shadow zone, and a semi-shadow zone. In the illuminated zone, the transmitting and receiving aerials are within the horizon distance, while in the shadow zone they are a greater distance apart. The centre of the semi-shadow zone is within the line-of-sight range [Eq. (2.70)]. The semi-shadow zone (the double-shaded area in Fig. 2.45) separates the other two zones.

In a first approximation, it is assumed that the boundaries of the semi-shadow zone are $0.8r_0$ and $1.2r_0$ distant from the transmitting aerial. More accurately, the illuminated zone is that which permits

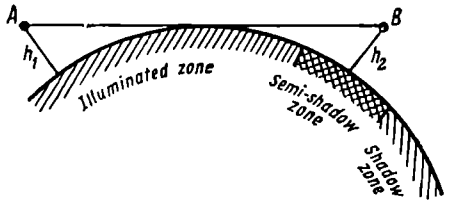


Fig. 2.45. Illuminated zone, shadow zone, and semi-shadow zone with radio waves diffracted around the earth's surface

the use of the interference equations. Within the shadow zone, use may only be made of the so-called one-term diffraction equations (derived on page 104). In other words, the outer boundary of the semi-shadow zone is determined by the validity of one-term equations. It is understood that radio waves find their way into the shadow zone owing to diffraction found around the earth.

The classical approach to the spherical-earth problem is to solve Maxwell's equations written for air and earth, with an appropriate choice of the boundary conditions, and to allow for the specific conditions at the radiator. Exactly this approach was used by the early investigators of the problem. Its limitation was that the expression for the Hertzian vector was obtained as a slowly converging series of Legendre, Hankel, and Bessel functions.

In 1918, Watson of England devised a method for solving this series [25]. The transformation of Watson applied to a case where the transmitting and receiving aerials were immediately on the earth's surface, and the earth was assumed to have an infinitely large conductivity. More specifically, Watson's solution held for wavelengths greater than 5000 metres, propagated over sea.

In 1935-1937, Vvedensky, elaborating on Watson's transformation, formulated a theory to explain the diffraction of radio waves around the spherical earth having a finite conductivity, with elevated transmitting and receiving aerials [26]. A major advantage of his theory was that it covered the USW band. Later (in 1942), Vvedensky

proposed a very convenient method for calculating the diffra- field in the USW band.

In 1937-1939, van der Pol and Bremmer of the Netherlands somewhat modified the asymptotic expansion of the func used in Vvedensky's theory. While Vvedensky determined poles of the integrand from an expression whose left-hand side the tangent of a complex argument (which is the reason why is called the "tangent approximation"), van der Pol and Brem found the poles from an expression whose left-hand side was a Ha function of order one-third (the so-called *Hankel approximat* However, the difference between the two approximations is small to have any practical significance.

The diffraction equations examined so far have the property as the distance is increased (within the shadow zone) the con gence of the series rapidly increases. This makes it possible to only one term in the expansion in many cases. Such formulae sometimes called one-term diffraction equations. Conversely, convergence of the series rapidly diminishes on moving towards boundary of the line-of-sight region, and the usual diffract equations are unsuitable.

In 1945, Fock [28] applied a new method to the diffraction of ra waves around the spherical earth. He replaced the slowly convergi series for the Hertzian functions with a contour integral in a complex plane, of a different form, however, than that used Watson's method thereby doing away with the awkward Hank function of order one-third. The contour of this integral pas in the first and second quadrants. Using the "large paramete concept, selecting the main area of integration, and replacing t Hankel and Bessel functions within that area with their asymptot expansions in terms of the re-introduced Airy integral function Fock arrived at a closed expression for the attenuation functio applicable to any distance from the transmitter. Analysis of t solution showed that at small distances from the transmitter t expression would reduce to the ordinary interference equation while at large distances the solution would change into a one-ter equation. Later, tables of Airy functions were compiled under Fock supervision, thereby making his solution applicable to practi cal calculations [29].

In 1946, Leontovich and Fock in a joint paper [30] showed tha the diffraction of radio waves around the spherical earth could b solved by a relatively simple parabolic-equation method whic they had used earlier for radio propagation over a flat earth.

Fock's design equation for elevated transmitting and receivi aerials [31] follows. Much as in the flat-earth problem we used th "scale distance" s to express the numerical distance x, so Foc

... the "scale distance" L and "scale height" H (since he ... elevated aerials).

... scale distance is defined as

$$L = \left(\frac{\lambda a^2}{\pi}\right)^{1/3} \text{ m} \tag{2.83}$$

... the scale height as

$$H = \frac{1}{2}\left(\frac{a\lambda^2}{\pi^2}\right)^{1/3} \text{ m} \tag{2.84}$$

... a result, the path length r and the aerial heights h_1 and h_2 can ... expressed in dimensionless units such that

$$x = \frac{r}{L} \tag{2.85}$$

$$y_1 = h_1/H \quad \text{and} \quad y_2 = h_2/H \tag{2.85a}$$

... referred to as the *relative distance* and the *relative aerial ...*, respectively.

... Fock's theory, the attenuation function is defined as

$$F = 2\sqrt{\pi x}\left|\sum_{s=1}^{\infty} \frac{e^{ixt_s}}{t_s + q^2} \frac{h_2(t_s + y_1)}{h_2(t_s)} \frac{h_2(t_s + y_2)}{h_2(t_s)}\right| \tag{2.86}$$

... is, the absolute value of the summation of the series.

In Eq. (2.86), q is a parameter describing the earth as a partially ...ducting surface and is defined as

$$q = i\sqrt[3]{\frac{\pi a}{\lambda}} \frac{1}{\sqrt{\varepsilon' - i60\lambda\sigma}} \tag{2.87}$$

Before going any further, it should be noted that for long waves ($\lambda \to \infty$) and a well conducting surface ($\sigma \to \infty$) the parameter q ...ends to zero. Conversely, for ultra-short waves ($\lambda \to 0$) and a ...ly conducting surface ($\sigma \to 0$), q tends to infinity.

In the same equation, $h_2(t)$ is an Airy integral function related ... the Hankel function of the second kind of order one-third as follows:

$$h_2(t) = \sqrt{\frac{\pi}{3}}\, e^{-i\frac{2\pi}{3}} t^{1/2} H^{(2)}_{1/3}\left[\frac{2}{3} t^{3/2}\right] \tag{2.88}$$

There are tables compiled for the Airy integral functions and their first derivatives for a complex argument [32]. Asymptotic expansions for the Airy integral functions are given in [28].

In Eq. (2.86), t_s are the roots of the equation

$$h_2'(t) - qh_2(t) = 0 \tag{2.89}$$

They are numbered in ascending order of magnitude. The values of the first five roots are given in Table 2.4 for two limiting values of q, namely for $q = \infty$ (the USW band) and $q = 0$ (very long waves).

Table 2.4. VALUES OF THE FIRST FIVE ROOTS, t_s, for $q=\infty$ and $q=0$

s	$q=\infty$ (USW)	$q=0$ (VLW)	s	$q=\infty$ (USW)	$q=0$ (VLW)
1	$2.338e^{i\frac{2\pi}{3}}$	$1.019e^{i\frac{2\pi}{3}}$	4	$6.787e^{i\frac{2\pi}{3}}$	$6.163e^{i\frac{2\pi}{3}}$
2	$4.088e^{i\frac{2\pi}{3}}$	$3.248e^{i\frac{2\pi}{3}}$	5	$7.994e^{i\frac{2\pi}{3}}$	$7.372e^{i\frac{2\pi}{3}}$
3	$5.521e^{i\frac{2\pi}{3}}$	$4.820e^{i\frac{2\pi}{3}}$			

Analysis of (2.86) shows that the series converges rapidly on moving farther into the shadow zone and that it suffices to leave only the first term for engineering calculations. This "one-term" equation has the form

$$F = 2\sqrt{\pi x}\left|\frac{e^{ixt_1}}{t_1+q^2}\right|\left|\frac{h_2(t_1+y_1)}{h_2(t_1)}\right|\left|\frac{h_2(t_1+y_2)}{h_2(t_1)}\right| \qquad (2.90)$$

For the semi-shadow zone, use is made of several early terms in the series.

The one-term diffraction equation may be presented as the product of three terms

$$F = U(x)\,V(y_1)\,V(y_2) \qquad (2.91)$$

the first of which depends on distance, and the second and third on the heights of the transmitting and receiving aerials, respectively. The last two terms are referred to as height factors. They have an identical structure and enter the diffraction equation in a perfectly symmetrical way. This guarantees that the reciprocity theorem is satisfied. In our case,

$$U(x) = 2\sqrt{\pi x}\left|\frac{e^{ixt_1}}{t_1+q^2}\right| \qquad (2.92)$$

$$V(y) = \left|\frac{h_2(t_1+y)}{h_2(t_1)}\right| \qquad (2.93)$$

If both aerials are placed on the ground, the height factors reduce to unity, and the diffraction equation takes a still simpler form

$$|F| = 2\sqrt{\pi x}\left|\frac{e^{ixt_1}}{t_1+q^2}\right| \qquad (2.94)$$

In [28] Fock shows that t_s for the limiting values of q can be found by any one of the following equations

if $\left|\frac{q}{\sqrt{t_s}}\right| < 1$,

$$t_{s,\,q} \cong t_{s,\,q=0} + \frac{1}{t_{s,\,q=0}}\,q \tag{2.95}$$

if $\left|\frac{q}{\sqrt{t_s}}\right| > 1$,

$$t_{s,\,q} \cong t_{s,\,q=\infty} + \frac{1}{q} \tag{2.96}$$

The calculation of the field strength in the shadow zone in the long, medium and most of the short-wave band is largely simplified owing to the fact that under the specified conditions the transmitting and receiving aerials may be regarded as placed immediately at the earth's surface. The attenuation function is found by Eq. (2.94) and is a function of the distance r, frequency f, and the soil characteristics ε' and σ.

Figs. 2.46 and 2.47 show two plots for the calculation of the ground-wave field in the wave band from 30,000 to 30 metres, the distances from 100 to 2000 kilometres, and two types of surface such that (1) $\varepsilon' = 80$; $\sigma = 4$ siemens per metre (sea water); and (2) $\varepsilon' = 4$, $\sigma = 0.01$ siemens per metre.

On short distances from the transmitter use has been made of the Mhuleikin-van der Pol equation, and on long distances, of the diffraction equation.

The plots hold for a radiated power of one kilowatt, with a vertical aerial located immediately at the earth's surface and having a gain of $G_1 = 1.5$. Exactly this type of aerial is usually employed on long and medium waves.

The plots of Figs. 2.46 and 2.47 have been borrowed from the recommendations of the 8th Plenary assembly of the International Consultative Radio Committee (CCIR), held in Warsaw in 1956 [33].

The plots are self-explanatory and easy to use. Noting that the field strength at the receiver is proportional to the square root of the radiated power, and writing $E_{rms\,1}$ for the field strength found from the plot for a radiated power of one kilowatt, we may write

$$E_{rms} \div E_{rms\,1} = \sqrt{P_{1(\mathrm{kW})}} \div 1$$

The field due to the transmitter is defined as

$$E_{rms} = E_{rms\,1}\sqrt{P_{1(\mathrm{kW})}} \tag{2.97}$$

If the parameters of the transmitter are specified in terms of the aerial height, the anti-node current, and the wavelength, the

transmitted power may be written as

$$P_1 = 160\pi^2 \left(\frac{h_{eq} I_{rms}}{\lambda}\right)^2 \text{ W} \qquad (2.98)$$

Example 2.12. Determine the field strength at a distance of 1200 kilometres from a transmitter operating on $\lambda = 1500$ metres and

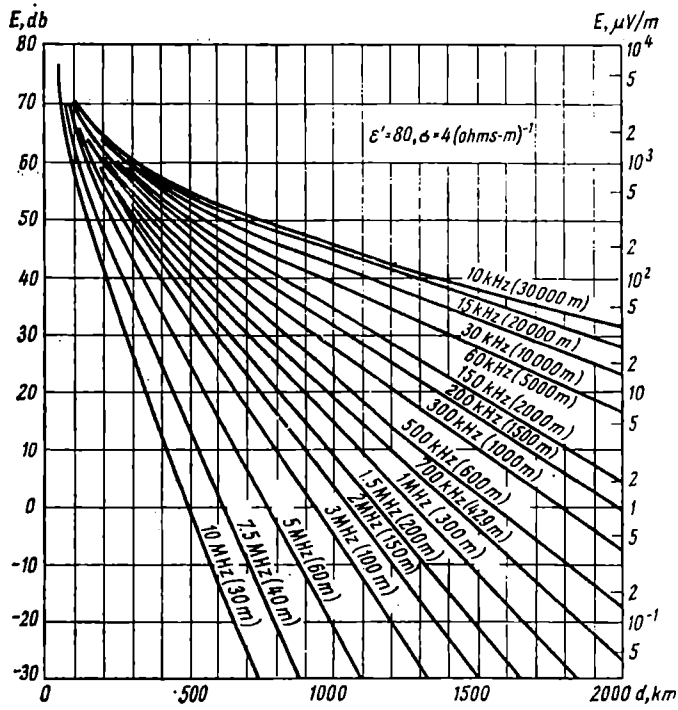


Fig. 2.46. Ground-wave field as a function of distance for propagation over seas; $\sigma = 4$ siemens/m, $\varepsilon' = 80$

radiating 200 kW. The waves are propagated over a surface of $\varepsilon' = 4$ and $\sigma = 0.01$ S/m.

Solution: From the plot of Fig. 2.47.

$$E_{rms\,1} = 10 \text{ mV/m}$$

Substituting it in Eq. (2.97) gives

$$E_{rms} = 10\sqrt{200} = 141 \text{ mV/m}$$

Example 2.13. Determine the transmitted power of a transmitter operating on $\lambda = 300$ m, that is necessary to produce a field of

E_{rms} = 40 mV/m at a distance of 600 km. The waves are propagated over a surface of $\varepsilon' = 80$ and $\sigma = 4$ S/m.

Solution: From the plot of Fig. 2.46

$$E_{rms\,1} = 60 \text{ mV/m}$$

substituting the values of E_{rms} and $E_{rms\,1}$ in Eq. (2.97) gives

$$P_1 = (E_{rms}/E_{rms\,1})^2 = 0.44 \text{ kW}$$

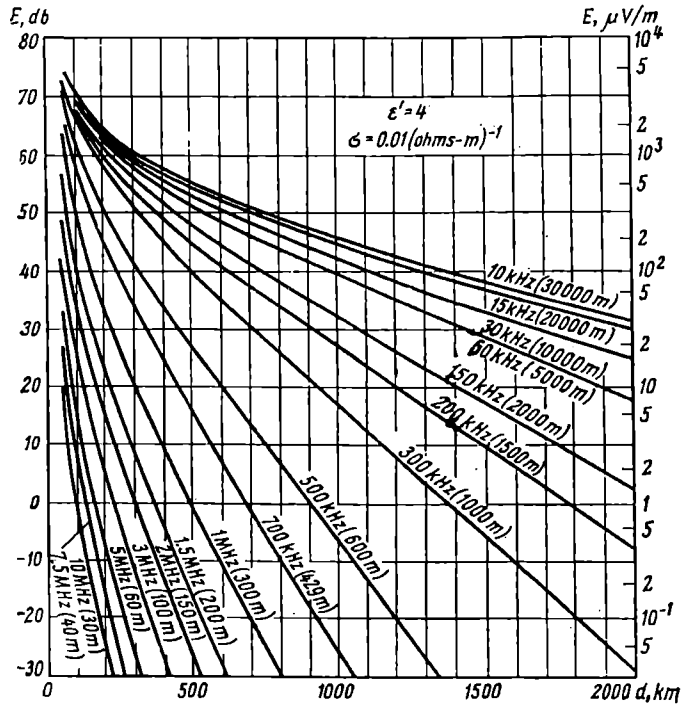


Fig. 2.47. Ground-wave field as a function of distance for propagation over land; $\sigma = 0.01$ siemens/m; $\varepsilon' = 4$

The diffracted field is much more difficult to determine in the UHW band, since in that band use is ordinarily made of elevated aerials, and one has to introduce additionally the heights of the transmitting and receiving aerials. Therefore, when the field at the point of reception must be found accurately and with allowance for the poor conductivity of the earth's surface, preference should be given to the complete diffraction equation (2.90).

In many practical cases involving microwave frequencies, it will suffice to calculate the field strength approximately, letting the wavelength tend to zero and q to infinity, that is, assuming that the parameter q is sufficiently large for any kind of surface. The orders of magnitude for q for vertically and horizontally polarized waves and four basic types of surface are given in Table 2.5.

Table 2.5. VALUES OF q AT MICROWAVE FREQUENCIES FOR DIFFERENT TYPES OF SURFACE

Wavelength	Vertical polarization				Horizontal polarization			
	Sea water	Fresh water	Moist soil	Dry soil	Sea water	Fresh water	Moist soil	Dry soil
10 metres	-0.6	$i15.4$	-9.7	-3	5280	9.2	129	2.4
	$+i2.3$		$+i37.1$	$+i59.2$	$+i3745$	$+i1230$	$+i435$	$+i241$
1 metre	-10.9	$i33.2$	-2.7	$i129$	2780	$i2658$	27	$i517$
	$+i15.2$		$+i90$		$+i3837$		$+i900$	
10 centimetres	-10.3	$i71.6$	$i193$	$i278.3$	845	$i5730$	$i1936$	$i1113$
	$+i69.5$				$+i5800$			
1 centimetre	$i154.6$	$i154.6$	$i417$	$i602$	$i12370$	$i12370$	$i4170$	$i2408$

As follows from Table 2.5, this technique can be applied to horizontal polarization at wavelengths under 10 metres, and to vertical polarization, at wavelengths not over 1 metre.

With q tending to infinity, the first root of Eq. (2.89), as follows from Table 2.4, is

$$t_{1\infty} = 2.34e^{i\frac{2\pi}{3}} = -1.17 + i2.02$$

Using (2.89), $h_2(t)$ may be written in terms of its first derivative. Substituting it in diffraction equation (2.90) gives

$$|F| = 2\sqrt{\pi x}\,\frac{e^{-x\,Im\,t_1}}{\left|1+\frac{t_1}{q^2}\right|}\left|\frac{h_2(t_1+y_1)}{h_2'(t_1)}\right|\left|\frac{h_2(t_1+y_2)}{h_2'(t_1)}\right| \qquad (2.99)$$

With q tending to infinity, the above equation takes the form

$$|F| \cong 2\sqrt{\pi x}e^{-2.02x}\left|\frac{h_2(t_1+y_1)}{h_2'(t_1)}\right|\left|\frac{h_2(t_1+y_2)}{h_2'(t_1)}\right| \qquad (2.100)$$

where the distance-dependent term and the height factors have values different from (2.92) and (2.93), quoted for the general

diffraction equation, namely

$$U_1(x) = 2\sqrt{\pi x}e^{-2.02x} \tag{2.101}$$

$$V_1(y) = \frac{h_2(t_1+y)}{h_2'(t_1)} \tag{2.102}$$

From (2.89) it follows that when q tends to infinity, the height factors tend to zero, while with $y = 0$, $V_1(0)$ tends to zero. Conse-

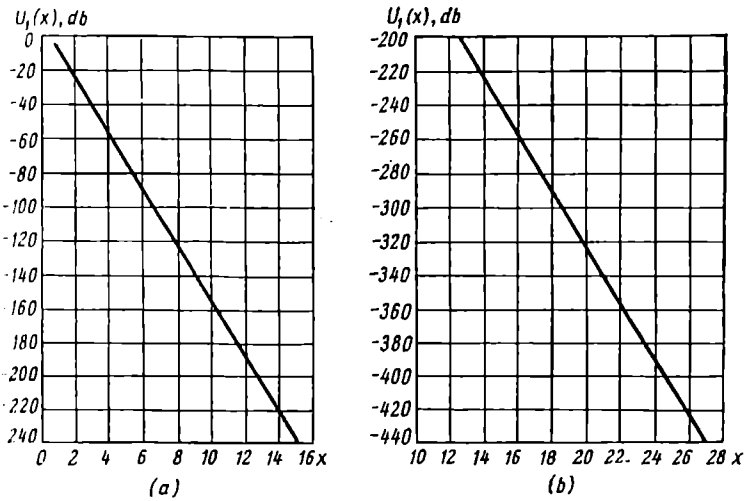


Fig. 2.48. A logarithmic plot of $U_1(x)$

quently, with the assumption made, the field at the earth surface turns to zero (even though the transmitting aerial is elevated).

Figs. 2.48 and 2.49 show plots (after [34]) of $U_1(x)$ and the height factors $V_1(y_1)$ and $V_1(y_2)$.

Then the attenuation function will be defined as

$$F = U_1(x)\, V_1(y_1)\, V_1(y_2) \tag{2.103}$$

which may be readily re-written logarithmically

$$F_{db} = U_1(x)_{db} + V_1(y_1)_{db} + V_1(y_2)_{db} \tag{2.104}$$

It is the values of $U_1(x)$, $V_1(y_1)$ and $V_1(y_2)$, expressed in decibels, that can be found from the plots of Figs. 2.48 and 2.49.

F in decibels can be converted to F as a proper fraction by use of the formula

$$F = 10^{\frac{F_{db}}{20}} \tag{2.105}$$

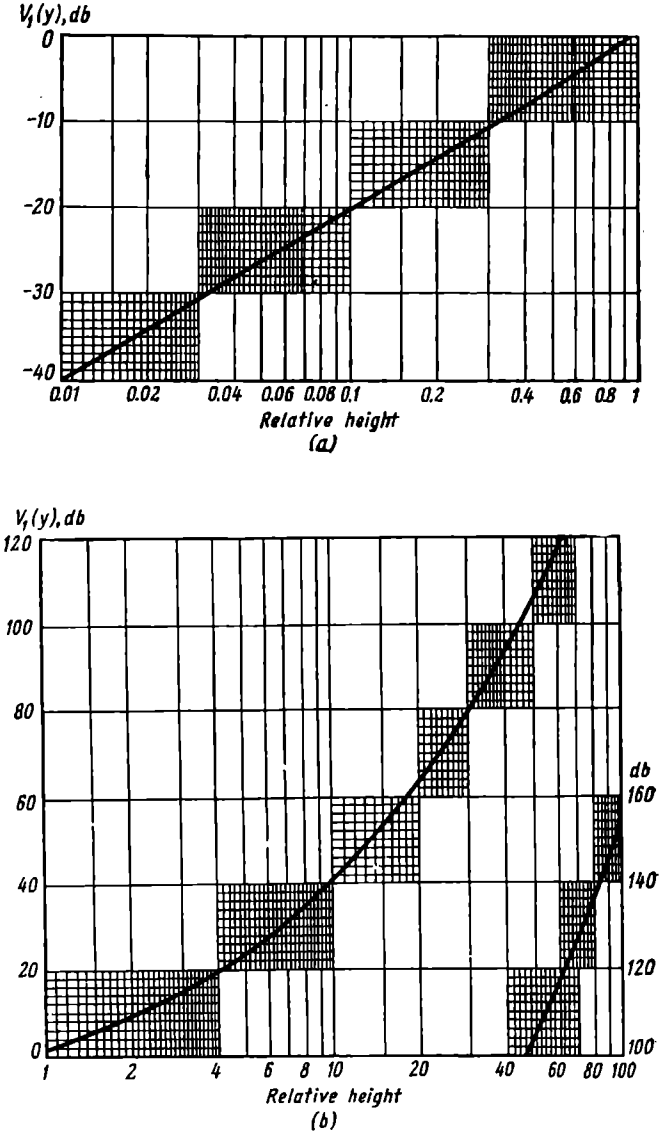


Fig. 2.49. A logarithmic plot of $V_1(y)$

As always, the field is given by

$$E_{rms} = \frac{173\sqrt{P_{1(\mathrm{kW})}G_1}}{r_{(\mathrm{km})}}F \qquad (2.106)$$

As follows from the foregoing, in contrast to the more rigorous method, the approximate one ignores the difference in electric characteristics between the various kinds of surface. Besides, it is assumed that the attenuation function is independent of the type of polarization. The latter assumption, however, holds only for wavelengths shorter than one metre.

Example 2.14. Determine the attenuation function and the received field using the plots of Figs. 2.48 and 2.49, for $P_1 = 25$ watts, $G_1 = 120$, $\lambda = 20$ cm, $h_1 = 25$ m, $h_2 = 10$ m, and $r = 35$ km.

Solution: To begin with, find the scale distance and the scale height from (2.83) and (2.84). Substituting $a = 6.37 \times 10^6$ metres in these equations gives

$$L = 1.37 \times 10^4 \text{ m}$$

and

$$H = 14.8 \text{ m}$$

Using (2.85) and (2.86), determine the relative distance x and the relative aerial heights y_1 and y_2:

$$x = \frac{3.5 \times 10^{-4}}{1.37 \times 10^4} = 2.56; \quad y_1 = \frac{25}{14.8} = 1.69;$$

$$y_2 = \frac{10}{14.8} = 0{,}670$$

Using the plots of Figs. 2.48 and 2.49, the values of $U_1(x)$, $V_1(y_1)$, and $V_1(y_2)$ are

$$U_1(x) = -30 \text{ db}; \quad V_1(y_1) = +8 \text{ db}; \quad V_1(y_2) = -3 \text{ db}$$

By Eq. (2.104),

$$F_{\mathrm{db}} = -30 + 8 - 3 = -25 \text{ db}$$

which, according to (2.105), corresponds to $F = 1 \div 17.8$.

Substituting the specified and found values in (2.106) gives

$$E_{rms} = \frac{173\sqrt{25 \times 10^{-3} \times 120}}{35 \times 17.8} = 0{,}48 \text{ mV/m}$$

In cases where an approximate value of the field strength will suffice, it may be of advantage to determine the attenuation function (and, consequently, the field strength) at the line-of-sight range, that is, at a distance where the interference equations will yield a field of value zero. This method applies to horizontally polarized waves; with vertical polarization it yields less accurate results

and may only be used for wavelengths shorter than 50 centimetres [24].

The attenuation function at the line-of-sight range is given by

$$F = \frac{r_0}{L} F(y_1, y_2) \tag{2.107}$$

where $F(y_1, y_2)$ is taken from the plot of Fig. 2.50 in which the function is expressed in decibels. In order to convert it into a proper

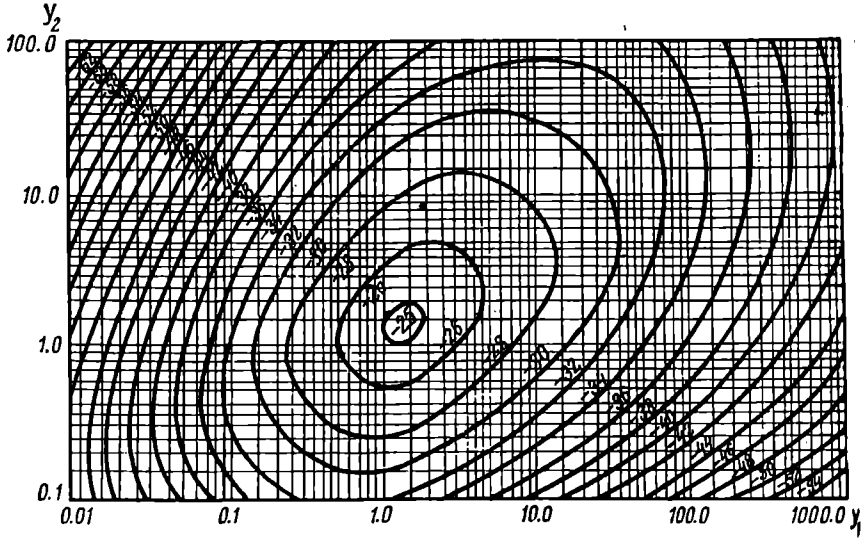


Fig. 2.50. A logarithmic plot of $F(y_1, y_2)$

fraction, use should be made of (2.105). If F should preferably be expressed in decibels (such as in determining the transmission loss), Eq. (2.107) may conveniently be re-written as

$$F_{\text{db}} = 20 \log_{10} \frac{r_0}{L} + F(y_1, y_2)_{\text{db}} \tag{2.107a}$$

Example 2.15. Find the attenuation function and the field strength at the line-of-sight range for the conditions specified in Example 2.14, assuming that the aerial emits horizontally polarized waves.

Solution: The line-of-sight range is given by (2.70a):

$$r_0 = 3.57 (\sqrt{25} + \sqrt{10}) = 29.1 \text{ km}$$

The values of L, y_1 and y_2 are known from the previous problem:

$$L = 1.37 \times 10^4 = 13.7 \text{ km}; \quad y_1 = 1.69; \quad y_2 = 0.67$$

From the plot of Fig. 2.50,

$$F(y_1, y_2) = -26 \text{ db}$$

or

$$F(y_1, y_2) = 0.05$$

Substituting the found values in (2.107), the attenuation function is

$$F = \frac{29.1 \times 0.05}{13.7} = 0.106$$

Thus, the attenuation function at the line-of-sight range is twice as great as in the previous problem.

The received field can be found by the standard equation

$$E_{rms} = \frac{173\sqrt{25 \times 10^{-3} \times 120}}{29.1} \times 0.106 \cong 1.1 \text{ mV/m}$$

To conclude our discussion of the spherical-earth problem, we shall consider a numerical example often encountered in practice.

Example 2.16. Calculate and plot the field strength as a function of distance for $P_1 = 50$ W, $G_1 = 100$, $h_1 = 60$ m, $\lambda = 10$ cm, $h_2 = 20$ m. The wave is vertically polarized and propagated over moist soil of $\varepsilon' = 10$ and $\sigma = 0.01$ S/m. The propagation is under conditions of standard atmospheric refraction, which fact is allowed for by giving the earth's radius its equivalent value $a_e = 8.5 \times 10^6$ m (see Sec. 3.4 below). Carry out calculations in the interval between the third maximum (in the direction of decreasing distance) and distance $2r_0$, where r_0 is the line-of-sight range.

Solution: Obviously, the first step should be to find the line-of-sight range under conditions of standard atmospheric refraction. Substituting $a_e = 8.5 \times 10^6$ m for $a = 6.37 \times 10^6$ m in (2.70), we get

$$r_0 = 4.12\,(\sqrt{h_{1(m)}} + \sqrt{h_{2(m)}}) = 50.3 \text{ km}$$

The illuminated region may be taken to extend out to $r' = 0.8r_0 \cong 40$ kilometres. The shadow region begins at $r'' = 1.2r_0 \cong 60$ kilometres.

The next step is to locate maxima and minima in the illuminated region, that is, to find the distances where $\cos\left(\theta^\circ + \frac{4\pi h_1' h_2'}{\lambda r}\right)$ in Eq. (2.9a) is $+1$ (at a maximum) and -1 (at a minimum). The reader will have noted that the argument of the cosine contains the reduced aerial heights so as to allow for the curvature of the earth's surface.

Since the maxima and minima of the field occur at small distances from the transmitter, the earth's surface may, to a first approximation, be taken as flat. Now, assuming that $\theta \cong 180°$, the appro-

ximate positions of maxima and minima may be found by (2.17a) and (2.18a). Substituting the values of h_1, h_2, and λ gives the values presented in the second column of Table 2.6.

Table 2.6. DISTANCES TO MINIMA AND MAXIMA FOR A FLAT AND A SPHERICAL EARTH

Extremum	Distance for flat earth, km	Actual distance for spherical earth, km
1st maximum	48	25
1st minimum	24	19
2nd maximum	16	14.5
2nd minimum	12	12
3rd maximum	9.6	9
3rd minimum	8	—

Thus, the minimum distance in our calculations should be 6 kilometres, and the maximum one, 3 $r_0 = 150$ kilometres.

The exact location of maxima and minima (that is, with allowance for h_1', h_2', and θ) and the actual values of the reflection coefficient R can be determined by a graphical method.

Choosing arbitrary distances in the interval between 8 and 40 kilometres, plot $\left(\theta^\circ + \frac{4\pi h_1' h_2'}{\lambda r}\right)$ and the reflection coefficient R as functions of distance. For this purpose, it will be convenient to arrange the calculated values as shown in Table 2.7.

Table 2.7

r, km	p	m	$h_1'h_2'$, m²	n	γ (rad)	θ°	R	$\left(\theta^\circ + \frac{720 h_1' h_2'}{\lambda r}\right)$
8	0.25	0.9	1080	0.96	0.01	180	0.93	1140
16	0.50	0.83	960	0.88	0.004	180	0.97	628
24	0.75	0.60	720	0.74	0.002	180	0.98	396
40	1.25	0.14	160	0.34	0.0007	180	0.99	211

In Table 2.7, m is taken from the plot of Fig. 2.43, $h_1'h_2'$ is found by (2.79), n from the plot of Fig. 2.44, γ by Eq. (2.81), 0° and R from Fig. 1.16 for vertical polarization, and $p = \frac{r}{\sqrt{2a_e h_1}}$ is a parameter necessary to determine m and n from Figs. 2.43 and 2.44.

The values of $\alpha = \left(\theta^\circ + \frac{720h_1'h_2'}{\lambda r}\right)$ and R as functions of distance are plotted in Fig. 2.51. The field is a maximum at $\alpha = 360°$, 720°, and 1080°, and a minimum at $\alpha = 540°$, 900°, and 1260°. At the maxima, the field strength is

$$E_{rms} = \frac{173\sqrt{P_{1(\mathrm{kW})}G_1}}{r_{(\mathrm{km})}}(1+R)\ \mathrm{mV/m}$$

and at the minima,

$$E_{rms} = \frac{173\sqrt{P_{1(\mathrm{kW})}G_1}}{r_{(\mathrm{km})}}(1-R)\ \mathrm{mV/m}$$

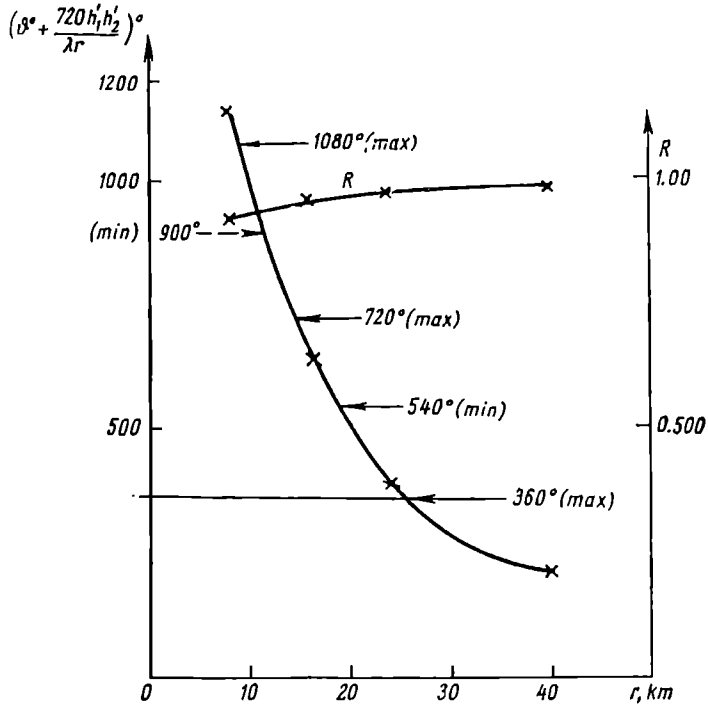


Fig. 2.51. A plot of $(\theta^\circ + \frac{720h_1'h_2'}{\lambda r})$ and R as functions of distance

The distances at which α takes values corresponding to an extremum of the attenuation function (or, which is the same, an extremum of E_{rms}) are listed in the last column of Table 2.6. At short distances from the transmitter, they are the same as for a flat earth. The design

values of the field strength at extremal points are summarized in Table 2.8.

Table 2.8. DESIGN VALUES OF FIELD STRENGTH AT EXTREMAL POINTS

r, km	$E_{rms\,(max)}$, mV/m	$E_{rms\,(min)}$, mV/m
9	83	
12		1.9
14.5	52	
19		0.6
25	31	

The field at a distance of 40 kilometres is found by Vvedensky's equation, (2.23), and the field in the shadow zone at distances of 80, 120, and 180 kilometres, by simplified diffraction equation (2.103). The results are tabulated in Table 2.9.

Table 2.9. RESULTS OF CALCULATION

r, km	m	x	$U_1(x)$, db	$V_1(y_1)$, db	$V_1(y_2)$, db	F, db	E_{rms}, mV/m
40	0.14	—	—	—	—	—	5.2
80	—	6	−90	+28	+10	−52	0.01
120	—	9	−140	+28	+10	−102	3×10^{-5}

By Eq. (2.83), the scale distance is

$$L=\left(\frac{\lambda a_e}{\pi}\right)^{\frac{1}{3}}=1.32\times10^4\text{ m}=13.2\text{ km}$$

and the scale height

$$H=\frac{1}{2}\left(\frac{a_e\lambda^2}{\pi^2}\right)^{\frac{1}{3}}=10\text{ m}$$

Tables 2.8 and 2.9 are shown graphically in the plot of Fig. 2.52. The distances are laid off as abscissa on a linear scale, and the field strength, as ordinate on a logarithmic scale. As is seen, the curves for the illuminated region match the design points in the shadow region well. The broken vertical line represents the line-of-sight range.

Radio Propagation over an Inhomogeneous Spherical Earth

Expanding their work on the propagation of ground waves over an inhomogeneous flat earth, Kalinin and Feinberg investigated in 1957 the more general case of radio propagation over an inhomo-

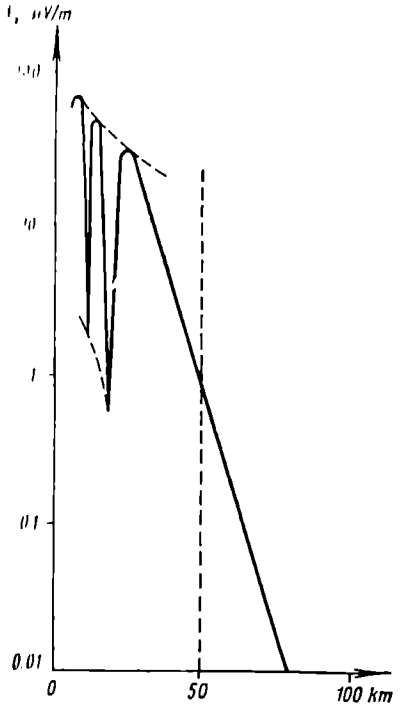


Fig. 2.52. Field strength as a function of distance in Example 2.16

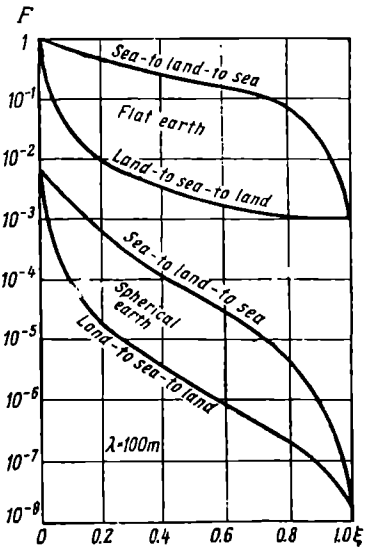


Fig. 2.53. Attenuation function versus the "land-fill" factor for spherical-earth propagation ($\lambda = 100$ m)

geneous spherical earth [35]. What follows is a brief outline of their findings.

An analysis of their solution shows that radio propagation over an inhomogeneous spherical earth is markedly affected by the intervening sections of the path. The absorption of wave energy and phase shift occurring along the path accumulate almost additively and are transferred to the point of reception. While over a flat earth radio waves are mainly propagated at a certain height above the ground and experience marked attenuation solely

within the "take-off" and "touch-down" terminals (Fig. 2.34), over a spherical earth the radio waves graze at the surface. The terminal sections still play an important part in the process of propagation, but their effect is not so decisive as in the flat-earth problem.

The effect of the earth's curvature stands out clearly from reference to the plots of Figs. 2.53 and 2.54. In effect, these plots are an extension of the relationship presented in Fig. 2.33. In these plots, the land-fill factor is laid off as abscissa, and the attenuation function, as ordinate. The plot of Fig. 2.53 applies to a wavelength of 100 metres, and that of Fig. 2.54, to a wavelength of

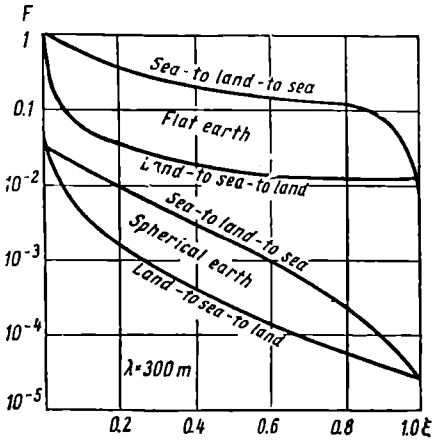


Fig. 2.54. Attenuation function versus the "land-fill" factor for spherical-earth propagation ($\lambda = 300$ m)

300 metres. In calculations, it was assumed that sea water has a conductivity of 4 siemens per metre, and land, 9×10^{-3} siemens per metre. For convenience, the respective curves for a flat earth are plotted in the same graphs.

A comparison of the curves for the flat and spherical earth clearly shows that in the latter case the intervening sections have a greater effect than in the former. The "sea-to land-to sea" and "land-to sea-to land" curves for a spherical earth, especially in Fig. 2.53, approach the straight line connecting the ends of the curves.

2.10. Significant Volumes of Radio Transmission

Placed in free space, an isotropic aerial radiates energy in all directions. Consequently, the power flux density has a finite value at every point in the surrounding space (provided the distance from

the radiator is not very great). Consider how a radio wave reaches point B (Fig. 2.55) at a distance r from the source.

To begin with, we may assume that the energy of a radio wave is conveyed to point B by a thin thread-like ray, AB, with the ray defined in terms of geometrical optics. Alternatively, we may assume that the wave energy reaching point B occupies a certain finite volume in the space around the axis AB. Which of the two assumptions is true can be shown by experiment.

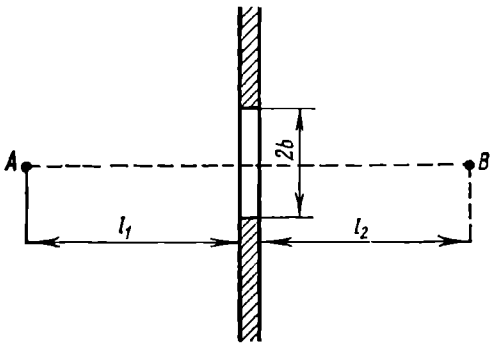


Fig. 2.55. Determining the volume significance to energy transmission

Let us put a radio-opaque diaphragm in the way of radio waves. Let, further, the diaphragm have an aperture of a variable diameter $2b$. This diaphragm may be likened to the iris diaphragm used in photographic cameras. To begin with, we shall leave the aperture widely open and measure the field strength at point B. Then we shall close down the aperture gradually until the meter shows a marked decrease in field strength. Noting the respective aperture setting, we may determine the diameter of the volume significant to energy transmission. Setting up the window at different distances l from the source, we can determine the boundaries of the significant volume.

If the assumption about energy transmission within a thread-like ray were true, we would note a sudden decrease in meter readings only when the aperture is fully closed.

Alternatively, the shape of the volume significant to radio transmission can be determined analytically, using Huygens' principle and the concept of Fresnel zones.

In its elementary form, Huygens' principle states that each point on the front of an advancing wave (produced by some primary source) is the source of a new, secondary spherical wave. Using Huygens' principle, it is possible to calculate the electromagnetic

field at any point in space from the known field strength on the wave-front surface.

Let the source of the primary wave be at point A (Fig. 2.56), and let S be a closed surface surrounding the source. In a special case, this may be an equi-phase spherical surface (that is, the surface of the wave front).

We inquire about the field due to the wave at point M (outside the closed surface) from the known value of the field on the surface S.

Let ψ denote the rectangular component of the sought field (electric or magnetic) at point M, ψ_S the same component on the surface S, and r the distance between the various points on the surface S and point M. According to Huygens' principle, the secondary field at point M due to a surface element dS is proportional to the field due to the primary wave at the element and the area of the surface element, that is

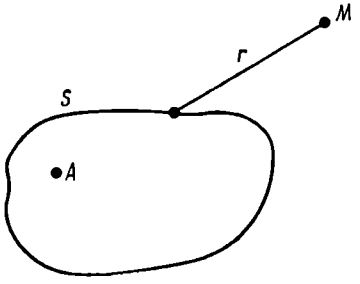


Fig. 2.56. Pertaining to Huygens' principle

$$d\psi = A\psi_S \frac{e^{-ikr}}{r} dS \qquad (2.108)$$

where A is the proportionality coefficient, and e^{-ikr} describes the natural relationship between the field phase and distance for a spherical wave [see Eq. (1.14a)].

The total field at point M is determined by adding together the elementary fields all over the surface S and is described by the integral

$$\psi = \int_S A\psi_S \frac{e^{-ikr}}{r} dS \qquad (2.109)$$

If the surface S is a plane, then, as is shown in texts on physics,

$$A = \frac{i}{\lambda} \cos(\mathbf{n}, \mathbf{r})$$

where $\mathbf{n}$ is an external normal to the plane and λ is the wavelength. Then Eq. (2.109) takes the form

$$\psi = \frac{i}{\lambda} \int_S \cos(\mathbf{n}, \mathbf{r}) \psi_S \frac{e^{-ikr}}{r} dS \qquad (2.110)$$

In a more general case, when the surface S is of an arbitrary shape, the mathematical statement of Huygens' principle takes

the form known as Kirchhoff's equation

$$\psi = -\frac{1}{4\pi}\int_S \left[\psi_S \frac{\partial}{\partial n}\left(\frac{e^{-ikr}}{r}\right) - \frac{e^{-ikr}}{r}\frac{\partial \psi_S}{\partial n}\right] dS \qquad (2.111)$$

As follows from Eq. (2.111), in order to determine the field at point M in the general case, one must know not only the value of ψ_S on the surface S, but also the values of its derivative along the normal to the surface. It should be noted that the integral

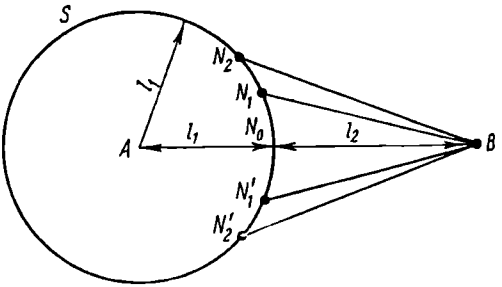


Fig. 2.57. Fresnel zones on the surface of a sphere

should be taken over the entire surface S, including its elements not "seen" from point M.

With Huygens' principle, it is possible to determine that part of the space which contributes significantly to radio propagation. For this purpose, let us draw a sphere of radius l_1 from point A as its centre and let it represent the surface of the wave front (Fig. 2.57). Knowing the values of ψ_S and $\frac{\partial \psi_S}{\partial n}$ on the surface S, we can calculate the field at point B by (2.111) and see that the result fully checks with the one given by (1.12).

In 1818, Fresnel of France showed that the construction appearing in Fig. 2.57 can be interpreted in a very obvious manner. Let l_2 denote the distance along the ray path AB, as measured from the point of reception to the surface of the sphere. We draw a set of curves from point B so that they cross the sphere S at points $(l_2 + \lambda/2)$ distant from point B.This set of curves forms a conical surface intersecting the plane of the drawing along the straight lines BN_1 and BN_1'. Similarly, the higher orders of conical surfaces are constructed, for which $BN_2 = l_2 + 2\frac{\lambda}{2}$ and, generally, $BN_n = l_2 + n\frac{\lambda}{2}$. The intersections between the conical surfaces and the sphere form a system of concentric circles. In Fig. 2.58 these circles are shown as seen from point B. The segments bounded by

adjacent circles have come to be known as *Fresnel zones*. The first Fresnel zone is a portion of the sphere bounded by a circle, while the higher-order zones are circular portions on the surface of the sphere.

The imaginary, or *virtual*, sources of secondary waves, located within the first Fresnel zone, are such that the difference in phase between the secondary waves at point B and the wave due to the virtual emitter at point N_0 does not exceed 180 degrees, because a path-length difference of a half-wavelength corresponds to a phase difference of 180 degrees. The waves produced by the virtual sources inside the second Fresnel zone differ in phase from the waves due to the emitter at N_0 by an angle from 180 to 360 degrees. It may be said that on the whole the waves due to the second Fresnel zone are 180 degrees out of phase with the waves due to the first zone. In Fig. 2.58, this difference is marked by assigning pluses and minuses to the alternate zones.

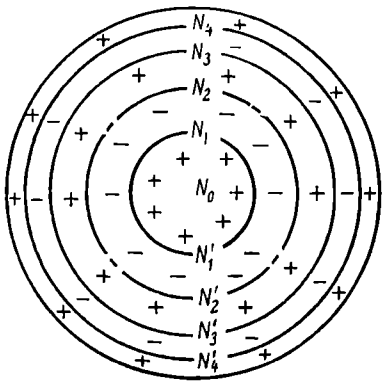


Fig. 2.58. Fresnel zones on the surface of a sphere

In texts on optics it is shown that adjacent higher-order zones cancel one another, this cancellation being the more complete, the higher the order of the zones. In the same texts it is shown that owing to this pair-wise cancellation the aggregate effect of all these zones is equal to about a half of that of the first zone. Thus, the first Fresnel zone bounds (and with a good margin at that) the volume contributing significantly to wave propagation.

As already noted, Fresnel zones may be constructed on surfaces of any arbitrary shape. It is convenient to choose for this purpose a plane normal to the ray path AB (Fig. 2.59). We inquire about the radius of the Fresnel zone for this case. By definition,

$$AN_n + N_nB = l_1 + l_2 + n\frac{\lambda}{2} \qquad (2.112)$$

From the triangles AN_nN_0 and BN_0N_n we have

$$AN_n = \sqrt{l_1^2 + b_n^2} \cong l_1 + \frac{b_n^2}{2l_1}$$

$$BN_n = \sqrt{l_2^2 + b_n^2} \cong l_2 + \frac{b_n^2}{2l_2}$$

because $b_n \ll l_1$ and $b_n \ll l_2$ always, so that we may use the binomial theorem. Substituting these quantities in (2.112) gives

$$\frac{b_n^2}{2}\left(\frac{1}{l_1}+\frac{1}{l_2}\right)=n\frac{\lambda}{2} \qquad (2.113)$$

whence

$$b_n=\sqrt{\frac{l_1 l_2 n\lambda}{l_1+l_2}} \qquad (2.114)$$

The radius of the first Fresnel zone is given by

$$b_1=\sqrt{\frac{l_1 l_2 \lambda}{l_1+l_2}}\ \text{m} \qquad (2.114a)$$

Points N_n are on the surface of an ellipsoid of revolution whose foci are at points A and B. It is this ellipsoid that contains the volume

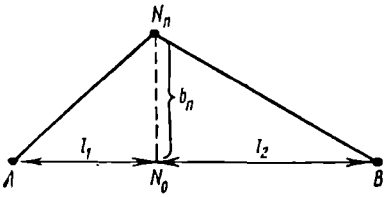


Fig. 2.59. Radii of Fresnel zones

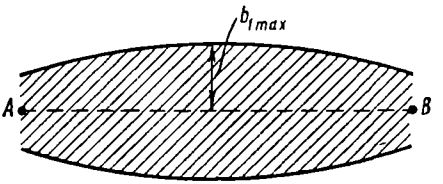


Fig. 2.60. The volume significant to wave propagation in free space

significant to radio propagation (Fig. 2.60). The radius of the first Fresnel zone is a maximum at the centre of the path.

Table 2.10 gives the radius of the first Fresnel zone at the centre of the path which is $l = l_1 + l_2 = 10$ kilometres long.

Table 2.10. RADIUS OF THE FIRST FRESNEL ZONE AT THE CENTRE OF THE PATH 10 KILOMETRES LONG FOR DIFFERENT WAVELENGTHS

λ, m	100	10	1	0.1	0.01	0.001
b_{max}, m	500	160	50	16	5	1.6

To sum up, the energy of a radio wave is transmitted within a certain definite volume having the shape of an ellipsoid of revolution and bounded by the first Fresnel zone, and not within a thread-like ray.

2.11. Line-of-Sight Radio Propagation over Hills

A terrain typical of most countries is one with hills. In such cases, the waves will run into obstacles, even though the transmitting and receiving aerials may be within the line-of-sight range. This is especially true of microwave radio links.

The point is that the size of an obstacle is a relative quantity, a function of frequency or wavelength. Thus, a terrain may be considered as a smooth surface for long and medium wavelengths, while the same terrain will cease to act as a smooth surface at ultrashort waves when hills a few tens of metres high present sizeable obstacles to propagation.

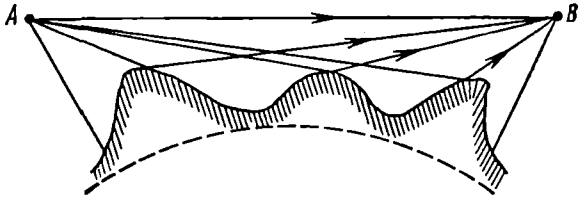


Fig. 2.61. Radio propagation over a hilly terrain

At first glance, it might appear that radio waves are propagated over hills in much the same way as they do over a smooth surface, except that instead of the single reflected ray which is formed over a smooth surface there might be several reflected rays coming from points where the angle of incidence is equal to the angle of reflection (Fig. 2.61). Actually, this is not so, because the reflected ray is formed within an area bounded by the first Fresnel zone and not at a geometrical point, and in most cases the tops of hills are much smaller in size than the first Fresnel zone.

An idea about the size of the first Fresnel zone in a reflecting plane may, in a first approximation, be obtained from the construction of Fig. 2.62. According to a well-known principle, the actual emitter A may be replaced with its image A'. Therefore, the reflected ray may be assumed to follow the path $A'B$. The shaded area in Fig. 2.62 represents the ellipsoid which is the boundary of the first Fresnel zone. At the point of intersection between the axis $A'B$ and the earth surface the diameter of the ellipsoid is $2b$. As will be recalled, the central section through an ellipsoid has the shape of an ellipse such that at small values of the grazing angle γ the dimension MN is many times the diameter $2b$.

Using a system of Cartesian coordinates in which the X-axis is directed along the major axis of the ellipse (Fig. 2.62) and limiting ourselves to aerials having the same height above a flat earth, when

the Fresnel zone boundary of radius b is a maximum at the reflection point, we may write the equation of an ellipse as follows

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \tag{2.115}$$

where a is the major axis of the ellipse approximately equal to $r/2$, and r is the distance between the transmitting and receiving aerials.

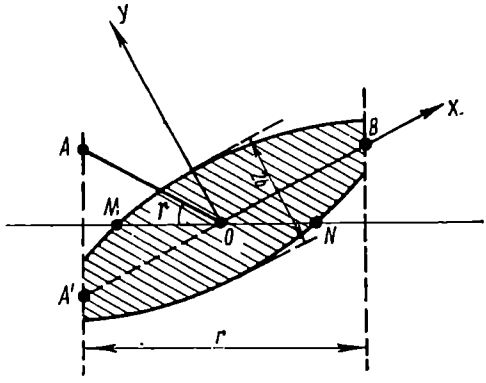


Fig. 2.62. Finding the size of the first Fresnel zone on a reflecting surface

The straight line MN can be described by an equation of the form

$$y = x \tan \gamma \tag{2.116}$$

whence the abscissa of the point of intersection between MN and the ellipse is

$$x=\frac{ab}{\sqrt{b^2+a^2\tan\gamma}}\ \text{m} \tag{2.117}$$

Since the angle γ is negligibly small, x may be taken to be indentically equal to the sought semi-axis ON of the ellipse on the earth surface.

The minor semi-axis of the ellipse, which is the boundary of the first Fresnel zone on the earth surface, is approximately equal to b.

As will be shown in the example that follows, the significant volume within which a reflected wave may be formed is rather large.

Example 2.17. Determine the area, bounded on the ground by the first Fresnel zone, where a reflected wave is formed, for $r = 50$ km, $h_1 = h_2 = 50$ m, and $\lambda = 10$ cm.

Solution: The radius of the first Fresnel zone for the ray $A'B$ at the reflection point is given by (2.114a)

$$b = \sqrt{\frac{l_1 l_2 \lambda}{l_1 + l_2}} \text{ m}$$

In our case,

$$l_1 = l_2 = r/2 = 25 \times 10^3 \text{ m},$$

whence

$$b = \sqrt{\frac{25 \times 10^3 \times 25 \times 10^3 \times 0.1}{50 \times 10^3}} = 35 \text{ m}$$

The grazing angle is given by (2.11)

$$\gamma = \frac{h_1 + h_2}{r} = \frac{50 + 50}{50 \times 10^3} = 2 \times 10^{-3}$$

Substituting the calculated and specified values in (2.117) and setting a approximately equal to $r/2 = 25 \times 10^3$ metres, because the grazing angle is negligibly small, we get

$$x = \frac{25 \times 10^3 \times 35}{\sqrt{35^2 + 25^2 \times 10^6 \times 4 \times 10^{-6}}} \cong 17 \times 10^3 \text{ m} = 17 \text{ km}$$

In a transverse direction, the minor axis of the ellipse bounding the significant region is 35 metres.

From the above example it follows that a reflecting region on the earth surface must be fairly large in area. On the other hand, hills with a flat top tens of kilometres in diameter will occur seldom, if ever. This implies that in considering radio propagation over a hilly terrain one may as a rule ignore reflections from the surface.

Hills affect radio propagation in a different way, though. They jut out into the region where the waves are propagated, and attenuate the waves by screening off that region. Therefore, a task facing a microwave-link designer is to choose the sites for, and the heights of, the aerials in such a way that the significant region will pass clear of the hill tops. Then the hills will not attenuate the wave field, that is, the attenuation function will be unity, and the field itself will be equal to one in free space.

Typical conditions for a microwave link in a hilly locality are shown in Fig. 2.63. Let the aerials of adjacent stations be located at points A and B on hill tops I and IV. The shading in Fig. 2.63 represents the propagation region bounded by the first Fresnel zone. In our case, the "critical" points along the link are hill tops II and III. The link designer must choose such heights for aerials A and B that the clearance of the direct ray above hill tops II and III will be not less than the radius of the first Fresnel zone at the location of the respective hills.

Denoting the distances between the hill tops as l_1, l_2, and l_3, and using Eq. (2.114a), we find the requisite clearances above critical points *II* and *III*

$$b_{II} = \sqrt{\frac{l_1 (l_2 + l_3)\,\lambda}{l_1 + l_2 + l_3}} \text{ m}$$

and

$$b_{III} = \sqrt{\frac{(l_1 + l_2)\, l_3 \lambda}{l_1 + l_2 + l_3}} \text{ m}$$

The aerial heights h_1 and h_2 on hills *I* and *IV* must be such that the clearance above hill tops *II* and *IV* will be greater than b_{II} and

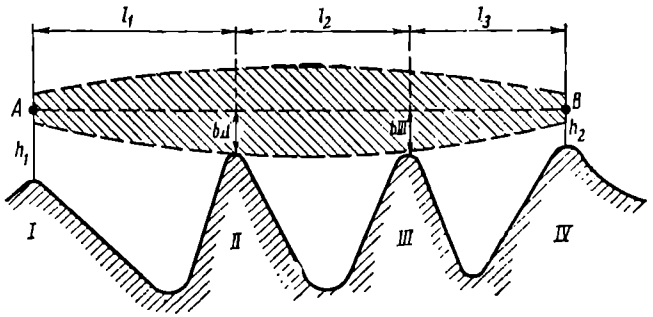


Fig. 2.63. Determining Fresnel zone clearances for a micro-wave link in a hilly locality, with the transmitting and receiving aerials set up within the line-of-sight range

b_{III}. That is, the first Fresnel zone clearance must be secured along the entire propagation path.

In practice, the Fresnel zone clearances have to allow for the curvature of the earth's surface. This complicates the matter, because the profile of the ground along the propagation path is usually specified only as elevations in so many metres above sea level. A way out is offered by a graphical method.

To make the profile in elevations more convenient for graphical construction, one has to exaggerate the height scale. The basis for the construction is provided by the equation for the horizon distance (2.69a), which may be re-written as

$$y = \frac{x^2_{(\text{km})}}{12.8} \text{ m} \qquad (2.118)$$

Eq. (2.118) in Cartesian coordinates is the equation of a parabola, where the distances (in kilometres) are laid off as abscissa, and the heights (in metres), as ordinate. On choosing a suitable scale for distances (in kilometres) and heights (in metres), we may construct

the profile of a flat earth by the equation

$$y = \frac{r^2_{(\mathrm{km})}}{4 \times 12.8} - \frac{\left(\frac{r_{(\mathrm{km})}}{2} - x_{(\mathrm{km})}\right)^2}{12.8} \text{ m} \qquad (2.119)$$

where r is the total length of the link.

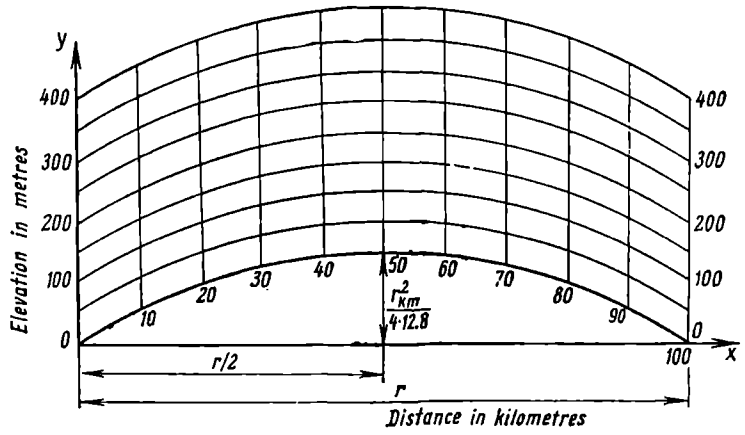


Fig. 2.64. The scale grid for a profile in elevations

The construction by Eq. (2.119) appears in Fig. 2.64. It should be remembered that the distances are laid off along the horizontal

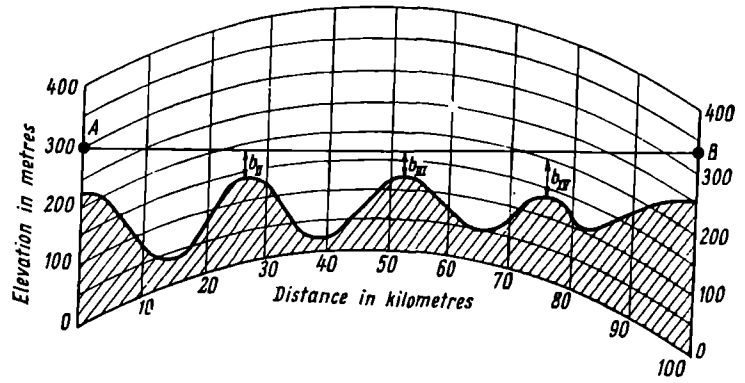


Fig. 2.65. Fresnel zone clearances found by means of an elevation profile plotted on a scale grid

axis, while the heights are laid off along the verticals and not along the normals to the earth surface, as would be customary.

The lines of equal heights follow the parabolic shape of the earth's surface and are spaced an equal distance apart. Using this grid, it is possible to plot the profile of a locality, calculate the requisite clearance above each hill top by Eq. (2.114a), lay off the clearances above the critical hills, and draw a straight line connecting the aerials at points A and B so that it will pass clear of all hill tops as required. The complete construction is shown in Fig. 2.65.

2.12. Line-of-Sight Radio Propagation over a Rough Surface

As with any obstacles, a rough surface is a relative term in radio propagation. Everything depends on the dimensions of rough spots on the earth surface in comparison with the wavelength used. At

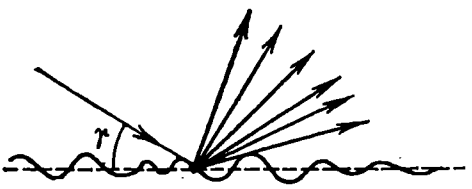


Fig. 2.66. Scattered reflections from a rough surface

very long wavelengths a terrain with hills rising about a hundred metres high may well be treated as a smooth surface; in the centimetric wave-band an even field covered with grass about ten centimetres high will be a rough surface.

In optics, the roughness of a reflecting surface is evaluated in terms of Rayleigh's criterion. This concept may successfully be utilized in radio propagation. The problem is stated as follows. Let there be a rough surface on which a radio wave is incident at a grazing angle γ. We inquire about the size of the rough spots, or irregularities, at which a diffuse reflection takes place instead of the specular one (Fig. 2.66). For simplicity, we assume that all irregularities are of the same height h. As an alternative, we may deal with an average height. In Fig. 2.67, the full line represents the lower limit, and the broken line, the upper limit of the surface irregularities.

Both the lower and upper planes will contribute to the reflected waves. The waves reflected from the lower plane will have to cover an additional path of length $BAC = 2h \sin \gamma$, so that the phase shift between rays I and II will be $\Delta\varphi = \frac{2\pi}{\lambda} 2h \sin \gamma = \frac{4\pi h}{\lambda} \sin \gamma$.

Let us agree to class (somewhat arbitrarily) reflections as diffuse ones and the earth surface as a rough one, if the phase shift $\Delta\varphi$ is greater than $\pi/2$.

From the inequality $\frac{4\pi h}{\lambda}\sin\gamma < \frac{\pi}{2}$, we obtain Rayleigh's criterion

$$h < \frac{\lambda}{8\sin\gamma}\ \text{m} \qquad (2.120)$$

that is, a criterion by which a surface satisfying it may be treated as a smooth one.

Eq. (2.120) shows that the height of irregularities at which reflections become diffuse depends not only on wavelength, but also

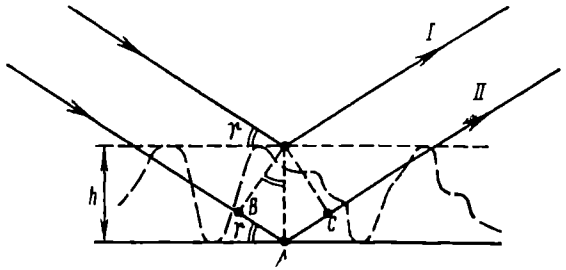


Fig. 2.67. Pertaining to the Rayleigh criterion.

on the grazing angle γ. As the grazing angle decreases, the height of obstacles must increase so as to cause diffuse reflections.

Rayleigh's criterion ignores the polarization of waves, although experimental data show that the type of polarization affects the conditions under which diffuse reflections can occur.

Since diffuse reflections entail the scattering of reflected-wave energy, the field strength in the direction of the reflected ray is decreased. Thus, irregularities on reflecting surfaces reduce the reflection coefficient.

To account for this reduction in the reflection coefficient in the case of ultra-short waves, elevated aerials and a rough reflecting surface, the interference equations for the attenuation function must be appropriately corrected. Besides, it is important to remember that a reflected ray is formed within an area bounded by the first Fresnel zone and not at a single point. It is within that area that the roughness of the reflecting surface must be evaluated.

At present, reliable theoretical methods for calculating the reflection coefficient for rough surfaces are still lacking, and so we shall limit ourselves to a few experimental data.

At wavelengths from 3 to 10 centimetres, most types of the earth's surface cause diffuse reflection. Therefore, the lobes of the radiation pattern of aerials in a vertical plane are not well defined. Specular reflections occur only from perfectly smooth surfaces (such as airfields, deserts, a quiet water surface, etc.).

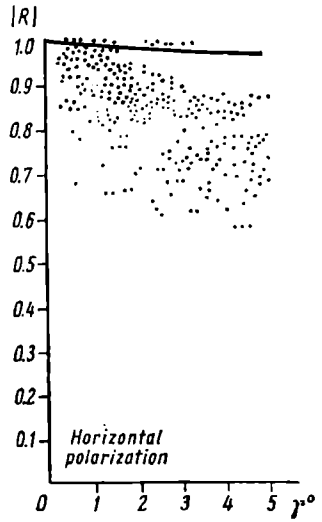


Fig. 2.68. Measured reflection coefficient as a function of grazing angle for horizontal polarization (sea, $\lambda = 10$ cm)

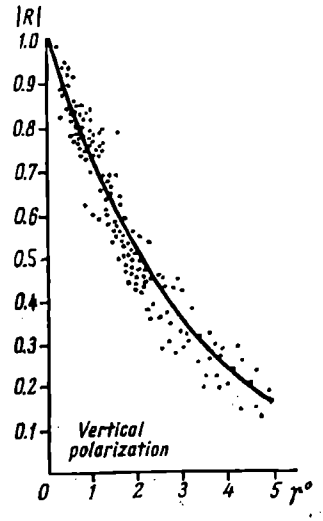


Fig. 2.69. Measured reflection coefficient as a function of grazing angle for vertical polarization (sea, $\lambda = 10$ cm)

At sea, the kind of reflection depends on the state of the sea, becoming diffuse when the sea is rough and also when the waves are horizontally polarized.

The effect of a rough sea on the reflection coefficient at $\lambda = 10$ centimetres is illustrated in Figs. 2.68 and 2.69. The grazing angles (those measured between the incident ray and the surface) are laid off as abscissa, and the respective reflection coefficients R, as ordinate. The dots represent the observed values, and the full lines, the theoretical values of the reflection coefficient.

The observed values of the reflection coefficient are closer to the theoretical ones with vertical polarization, which is a confirmation of the fact that surface irregularities affect the absolute value of the reflection coefficient more in the case of horizontal polarization.

Table 2.11 shows the effect of vegetation differing in height (1000 plants per square metre) for a wavelength of 9 centimetres, three elevation angles, and vertical and horizontal polarization.

Table 2.11. REFLECTED COEFFICIENTS FOR WAVES 9 cm LONG, REFLECTED FROM MOIST SOIL COVERED WITH VEGETATION

Surface	Grass height, cm	Reflection coefficient					
		$\gamma = 22°$		$\gamma = 36.5°$		$\gamma = 46.5°$	
		Polarization					
		V.	H.	V.	H.	V.	H.
No vegetation	0	0.30	0.86	0.50	0.78	0.58	0.74
Scattered grass shooting through soil	3-4	0.40	0.50	0.44	0.55	0.47	0.56
Grass clusters at places	9-12	0.18	0.65	0.23	0.58	0.33	0.49
Scattered bare spots	20-25	0.06	0.32	0.10	0.39	0.17	0.41
No bare spots	35-40	0.04	0.19	0.05	0.26	0.11	0.28

As is seen from the table, vegetation progressively reduces the reflection coefficient as the grass grows taller. This effect is more pronounced with vertical polarization and with decreasing angles of elevation.

Surfaces capable of specular reflection are a rare occurrence. Therefore, one must be prepared to reckon with diffuse reflections of waves shorter than one metre incident at grazing angles greater than two or three degrees.

Example 2.18. Determine the attenuation function at the receiver with allowance for the reflection of waves from a rough surface under the conditions specified in Example 2.3, and assuming that the surface irregularities reduce the reflection coefficient to 0.65.

Solution: As follows from Example 2.3, the attenuation function for a smooth surface is 0.314. Since the reflection coefficient is reduced, the attenuation function should be found by (2.9a), the complete interference equation.

Assuming, as before, that the phase of the reflection coefficient is 180 degrees and substituting $R = 0.65$, $\theta = 180°$, $h_1 = 25$ m, $h_2 = 10$ m, $\lambda = 1$ m and $r = 10$ km in Eq. (2.9a), we get

$$F = \sqrt{1 - 2 \times 0.65 \cos \left(\frac{720 \times 25 \times 10}{1 \times 10^4}\right)^\circ + 0.65^2} = 0.43$$

that is, the field strength at the receiver somewhat increases.

Of particular interest is a special case of a "rough" surface — a large city with its buildings, streets, squares, boulevards, and water spaces.

To begin with, we shall consider the fields of TV and VHF-FM radio broadcasts, when the transmitting and receiving aerials are within the line of sight.

Diagrammatically, the propagation of VHF waves under these conditions is shown in Fig. 2.70. As in propagation over a smooth surface, a direct and a reflected ray reach the receiving aerial. If the heights of the transmitting and receiving aerials are known, we may begin by finding the first Fresnel zone boundary. In the longitudinal direction this boundary is labeled by the letters MN in

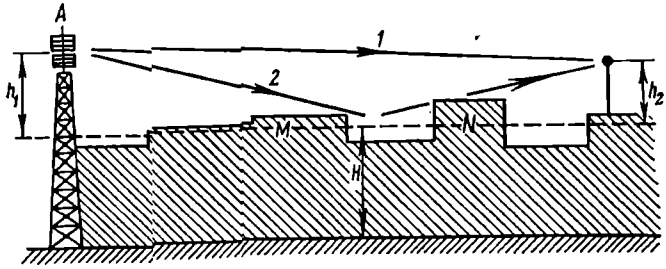


Fig. 2.70. VHF propagation within the line of sight in a city

Fig. 2.70. If this zone is within a built-up area, it is necessary to estimate the average roof-top height H there. Then the field may be calculated by use of the interference equations, with the aerial heights counted from the average level thus found. This assumption is a legitimate one, because at small grazing angles metal roofs are good reflectors.

2.13. Radio Propagation over a Knife Edge

In this Section, by "knife edge" is meant a sharply defined obstacle opaque to, and placed in the way of, radio waves. It is obvious that a smooth spherical earth cannot be classed as a knife-edge obstacle.

With such an idealized diffracting edge, devoid of any electrical properties, the diffracted field at the receiver may be calculated by methods widely used in physical optics.

Fig. 2.71 shows two cases of radio propagation across a knife-edge. At a, the obstacle does not cut the direct ray and only partly juts out into the significant volume. At b, the knife-edge cuts the direct ray AB. Let us agree to assign the quantity H a minus sign in the former case, and a plus sign in the latter.

In the theory of optical diffraction [36] it is shown that for free space the attenuation function is given by

$$F = \frac{1}{\sqrt{2}}[C(v) + iS(v)] = \sqrt{\frac{C^2(v) + S^2(v)}{2}}\, e^{i\nu} \qquad (2.121)$$

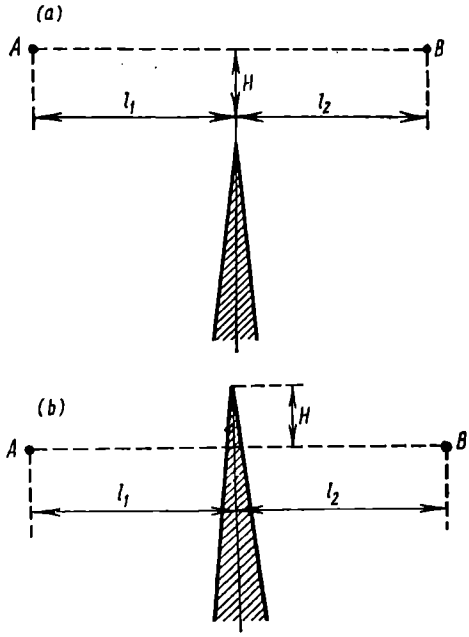


Fig. 2.71. Radio propagation across a knife-edge

where $C(v)$ and $S(v)$ are Fresnel integrals such that $-\frac{S(v)}{C(v)} = \tan\gamma$ and defined respectively as

$$\left.\begin{aligned} C(v) &= \frac{1}{2} - \int_0^v \cos\frac{\pi x^2}{2}\,dx \\ S(v) &= \frac{1}{2} - \int_0^v \sin\frac{\pi x^2}{2}\,dx \end{aligned}\right\} \tag{2.122}$$

Here the parameter v is given by

$$v = \frac{H\sqrt{2}}{b} \tag{2.123}$$

where b = radius of the first Fresnel zone at the obstacle;
H = height of the obstacle (which is positive at b and negative at a of Fig. 2.71).

The function $F(v)$ is plotted in Fig. 2.72. It should be noted that for $v > 2$ the attenuation function is approximated well by

$$-F(v)_{\text{db}} = 12.953 + 20\log_{10} v \tag{2.124}$$

In Fig. 2.72, the approximating line is shown dotted.

As Schelling, Burrows and Ferrel [37] showed in 1933, in the application of (2.121) to ultra-short waves it is important to remember that the knife-edge diffraction affects both the waves emitted by the transmitting aerial and the waves reflected from the earth

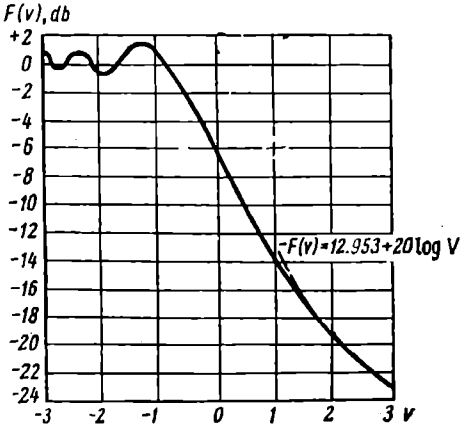


Fig. 2.72. Attenuation function versus the parameter v

surface on each side of the obstacle, that is, between the transmitting aerial and the obstacle, and between the obstacle and the receiving aerial. Fig. 2.73 shows that the field at point B is due to the combina-

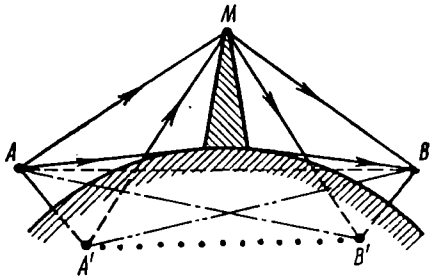


Fig. 2.73. The four-ray model of diffraction across a knife-edge

tion of four rays each of which has suffered refraction at the knife-edge. In the sketch, these rays are labelled AMB, $A'MB$, AMB' and $A'MB'$. The construction is based on the image principle according to which a reflected ray is treated as being due to an image of the actual emitter. Of course, the same reasoning applies to the receiving aerial. The addition of fields at the receiving aerial is

legitimate owing to the linearity of the entire system. For a long time this "four-ray" model of diffraction across a knife-edge obstacle was considered to be of purely academic interest and unrelated to the actual situation. In recent years, however, it has been found that the diffraction of radio waves at microwave frequencies is accompanied by a build-up of the field, or gain, past mountain ranges, which can satisfactorily be explained by the four-ray model of the propagation path.

It is important to realize what sort of gain it is. Let E_{dif} denote the field at point B in the absence of obstacle M, that is, with radio waves diffracted around a smooth spherical earth (it is assumed that point B is in the shadow region). It might intuitively appear that an obstacle, such as a tall mountain M, somewhere along the propagation path, would entail a further attenuation of the field. Facts show however that, given certain conditions, mountain ranges may be responsible for a field E_{ob} at point B to be in excess of the previously calculated field E_{dif}. That is, there is a gain in field strength as compared with the field set up due to the diffraction around a smooth earth.

There is a very simple physical explanation of this *obstacle gain*. As a rule, the obstacle M is much taller than the aerials of heights h_1 and h_2. Therefore, the attenuation factors F found for the four rays by Eq. (2.121) will be of the same order of magnitude. Besides, the reflection coefficient of earth in the USW band is close to unity. As a consequence, it may well happen that, given favourable phase relationships, the received field will be four times the field calculated by (2.121) for one ray. And this is the obstacle gain in the field beyond the obstacle.

In finding the resultant attenuation function for the four-ray propagation, it is important to take into account the phase relationships between the components. Let us use the following notation for the purpose:

(a) $n = 1$ will be the propagation path AMB; $n = 2$, $A'MB$; $n = 3$, AMB'; and $n = 4$, $A'MB'$;

(b) $G_n e^{i\psi_n}$ will be the attenuation for each of the four rays, found by (2.121), with the height H found relative to the respective base line, namely, AB for the first ray, and $A'B'$ for the fourth ray;

(c) $R_1 e^{-i\theta_1}$ will be the complex reflection coefficient along AM, and $R_2 e^{-i\theta_2}$, that along MB;

(d) $\psi_1 = \frac{2\pi}{\lambda}(A'M - AM)$ will be the phase shift due to the path length difference along the segment AM, and $\psi_2 = \frac{2\pi}{\lambda}(MB' - MB)$, that along the segment MB.

Then the attenuation function may be written as

$$F = |G_1 e^{i\gamma_1} + G_2 R_1 e^{i(\gamma_2 - \psi_1 - \theta_1)} + G_3 R_2 e^{i(\gamma_3 - \psi_2 - \theta_2)} + G_4 R_1 R_2 e^{i(\gamma_4 - \psi_1 - \psi_2 - \theta_1 - \theta_2)}| \quad (2.125)$$

If the obstacle height H is much greater than twice the aerial heights h_1 and h_2 (that is, the distances between the aerials and

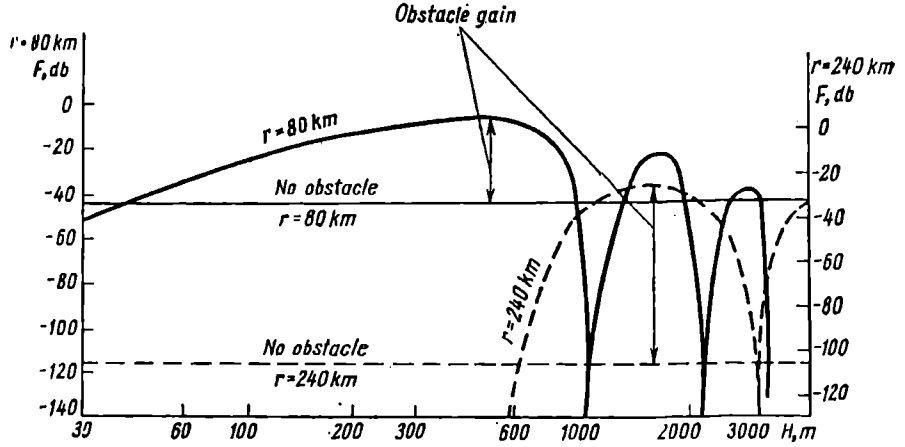


Fig. 2.74. Attenuation function versus the height of an obstacle, H

their images), Eq. (2.125) may be appreciably simplified. In such a case, we may set $G_n e^{i\gamma_n} = G_n e^{i\gamma}$, and neglect the phase factor $e^{i\gamma}$. Then,

$$F = G|1 + R_1 e^{-i(\psi_1 + \theta_1)}||1 + R_2 e^{-i(\psi_2 + \theta_2)}| \quad (2.125a)$$

Furthermore, if $R_1 = R_2 = 1$, and $\theta_1 = \theta_2 = \pi$, the equation finally reduces to

$$F = G|1 - e^{-i\psi_1}||1 - e^{-i\psi_2}| \quad (2.125b)$$

It is not difficult to see that if $\cos\psi_1 = -1$ and $\cos\psi_2 = -1$, the attenuation function takes the value $F = 4G$, as already noted.

Fig. 2.74, after [38], shows the attenuation function plotted against the obstacle height for propagation paths 80 kilometres (the full lines) and 240 kilometres (the broken lines) long, $h_1 = h_2 = 30$ metres, and a frequency of 100 megahertz. Over the 80-kilometre propagation path the maximum obstacle gain, corresponding to an obstacle height of 500 metres, is seen to be 23 decibels. On the 240-kilometre propagation path with an obstacle height of 1600 metres the obstacle gain is shown to be 80 decibels. In [38] is

also quoted the obstacle gain over a propagation path of 260 kilometres in Alaska. The mountain range, located approximately

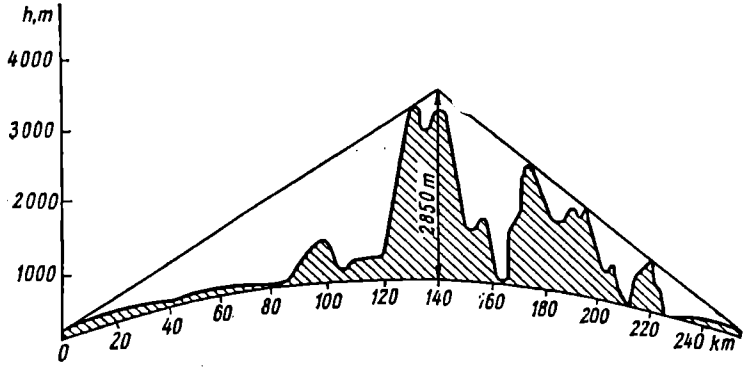


Fig. 2.75. Profile of a path subject to obstacle gain

mid-way between the transmitter and receiver, was 2500 metres high. The aerial heights were $h_1 = h_2 = 15$ metres. The elevation

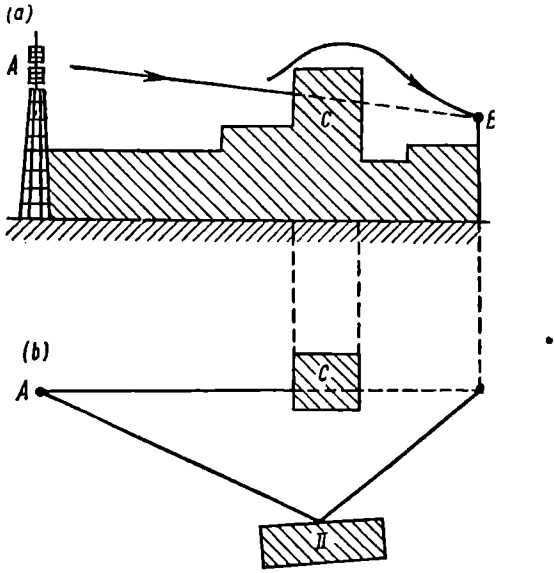


Fig. 2.76. Diffracting obstacles for ultra-short waves propagated in a city

profile of the propagation path is shown in Fig. 2.75. Measurements were taken at a frequency of 38 megahertz. The observed values

of the obstacle gain were a mere 10 decibels smaller than the calculated values.

In the case of TV and VHF-FM broadcasts in cities, individual buildings may well serve as diffracting obstacles. Such a case is diagrammatically shown in Fig. 2.76, where the diffracting obstacle is a building at C. Reception at point B is possible, firstly, because buildings may to some extent be transparent to ultra-short waves and, secondly, because the waves may diffract around the obstacle (Fig. 2.76*a*). In many cases, reception at point B is owing to the wave reflected from another building at D (Fig. 2.76*b*).

In view of the extremely complicated nature of USW propagation in cities, we are still lacking reliable analytical methods of calculation.

Chapter Three

The Effect of the Troposphere on Ground Waves. Propagation of Tropospheric Waves

3.1. General Properties of the Troposphere. Composition and Structure

The troposphere is the region of the atmosphere extending from the surface of the earth up to a height of 8 to 10 kilometres at polar latitudes, 10 to 12 kilometres at the moderate latitudes, and up to 16 or 18 kilometres at the equator.

In the troposphere, the percentage of the gas components does not vary with height, remaining practically the same as it is at the surface (see Table 4.1, page 210). The only exception is the water-vapour content which is strongly dependent on the weather conditions and sharply decreases with height.

The most important property of the troposphere is that its temperature decreases with height. The average vertical temperature gradient of the troposphere is 6 degrees per kilometre (about 5 degrees per kilometre in the lower troposphere and 7 degrees per kilometre in the upper troposphere). The annual average temperature of the air in the upper troposphere is —55° C in the polar regions, and —80° C at the equator. The upper boundary of the troposphere is called the tropopause which is a narrow region of constant temperature.

The cause of the gradual decrease in the temperature of air with height lies in the fact that the troposphere is almost transparent to sun rays which, on passing through it, heat it very little, if at all. The bulk of the solar energy is absorbed by the earth's surface. The heated earth surface is, in turn, a source of thermal radiation which heats the troposphere in an upward direction. Convective air movements also contribute to the heating of the troposphere — the air adjacent to the ground rises in temperature and moves upwards to give way to a colder air which is in turn heated and moves upwards, and so on. The non-uniform heating of ground areas produces ascending and descending air currents which result in turbulences and the mixing of air masses vertically, and this decides the temperature conditions in the troposphere.

Although the troposphere extends out to a relatively low height, it accounts for four-fifths of the entire air mass. The average pressure at the earth's surface is 1014 millibars (1 millibar = 1/1000 th bar. 1 bar = 10^5 newtons per square metre = 1.019 kilograms per square centimetre, which is very close to one atmosphere. 1 millimetre Hg = 1.332 millibars). At an altitude of 5 kilometres, it is nearly halved, at 11 kilometres it is 225 millibars, while at an altitude of 17 kilometres (the upper boundary of the troposphere at the equator), the atmospheric pressure is a mere 90 millibars.

The water vapour contained in the troposphere comes from the evaporation of water (likewise due to solar radiation) from the surface of the oceans, seas and water reservoirs. This is why the troposphere over an ocean carries more moisture than it does over a desert. The water-vapour content rapidly decreases with height. At an altitude of 1.5 kilometres, the water-vapour content is about one-half and at the upper boundary of the troposphere it is a few thousandths of what it is at the earth's surface.

The key characteristics of the troposphere are pressure p (in millibars), absolute temperature T (in degrees Kelvin, such that $T = t^\circ\,\mathrm{C} + 273$), and absolute humidity e (also in millibars). Sometimes, the water-vapour content of air is expressed in terms of specific humidity s (which is the mass of water vapour in grams per kilogram of air) or relative humidity r, expressed in per cent.

Absolute humidity e is related to s and r as follows

$$e = \frac{s_{\mathrm{g/kg}} p_{\mathrm{mbar}}}{623 - 0.377 s_{\mathrm{g/kg}}} \text{ mb} \tag{3.1}$$

and

$$e = \frac{E_{r\,\mathrm{mbar}} r}{100} \text{ mb} \tag{3.2}$$

where E_r is the pressure, found from charts, of water vapour which will saturate the air at a given temperature.

In 1925, the International Commission for Aeronavigation defined the so-called *international standard atmosphere* which is now often called the *standard troposphere*. This is a hypothetical troposphere having an arbitrarily selected set of characteristics reflecting an average condition of the real troposphere. The standard troposphere is considered to have a sea-level atmospheric pressure of approximately 1013 millibars, a sea-level temperature of 15° C, and a relative humidity of 60 per cent. It is further assumed that the pressure and temperature decrease upward at fixed rates, called the *standard pressure lapse rate* and the *standard temperature lapse rate*. The former is 12 millibars, and the latter 0.55 degree C, per 100 metres of height. Relative humidity remains unchanged at any height.

The standard troposphere is considered to extend out to an altitude of 11 kilometres. It is a convenient device in comparing the performance of aircraft, designing pressure altimeters, resolving problems in ballistics, and investigating radio propagation.

3.2. Refractive Index of the Troposphere

From the view-point of the refractive index, the troposphere may be treated as a mixture of dry air and water vapour. For each of these components the refractive index is a known quantity. The refractive index of the mixture can therefore be found from the known water-vapour content in the atmosphere, the partial pressure p_d of dry air, and the partial pressure e of water vapour, recalling that the refractivity of a mixture is the sum of the refractivities of the parts.

Since the refractive index n of the troposphere at the surface of the earth is only 0.0003 per cent greater than unity, it is more convenient to refer to variations in the refractive index in N units

$$N \equiv (n-1) \times 10^6 \tag{3.3}$$

where N is the *excess index of refraction*. It is the excess over unity of the refractive index, expressed in millionths. Thus, at the surface, where $n = 1.000325$, there is a value of 325 N units.

From physics it is known that the refractive index of a gas is

$$n = 1 + \rho (A + B/T) \tag{3.4}$$

where ρ is the gas density in kilograms per cubic metre, A is the constant (expressed in cubic metres per kilogram) dependent on the polarization of the gas molecules in an external field*, and B is the constant (expressed in cubic metres $\times$ degrees per kilogram) decided by the permanent molecular dipole moment.

Since, by Clapeyron's law, the density of a gas is directly proportional to its partial pressure and inversely proportional to its temperature in degrees absolute, Eq. (3.4) may be re-written as

$$N = \frac{C p_{part}}{T} (A + B/T) \tag{3.5}$$

where C is the proportionality factor, expressed in kilograms-degrees per cubic metres-millibars.

---

* The polarization of a molecule is the relative displacement of the positive and negative charges of the molecule, caused by an applied electric field. For a non-polar molecule, the electric dipole moment due to polarization is

$$m = \alpha E \text{ C/m}$$

where α is the polarizability of the molecule, in F-m^2, and E is the applied electric field, in volts per metre.

The gases making up dry air do not possess a permanent dipole moment. In contrast, the molecules of water vapour have a permanent dipole moment, $m = 6.13 \times 10^{-30}$ coulombs-metre. In view of the foregoing and noting that the total refractivity is the sum of the refractivities of the parts, the excess refractive index for moist air may be written as

$$N = \frac{C}{T} A_d p_d + \frac{Ce}{T} (A_w + B_w/T) \tag{3.6}$$

where A_d is the respective constant for dry air, while A_w and B_w are the constants for water vapour.

From experiments it is known that for dry air the product CA_d is 77.6 degrees per millibar. This seems to be true also of the product CA_w for water vapour. The ratio B_w/A_w has been measured fairly accurately and is assumed to be 4810. Substituting the numerical values of CA_d and B_w/A_w in Eq. (3.6) gives

$$N = \frac{CA_d}{T} \left(p_d + e + \frac{B_w e}{A_w T}\right) = \frac{77.6}{T_{deg}} \left(p_{\mathrm{mb}} + \frac{4810 e_{\mathrm{mb}}}{T_{deg}}\right) \tag{3.7}$$

where p is the total atmospheric pressure in millibars.

A thermodynamic approach to the processes occurring in the troposphere shows that in a well-mixed air and with the air temperature varying adiabatically* with height, the gradient of the excess refractive index within the troposphere remains constant

$$\frac{dN}{dh} = -4.45 \times 10^{-2} \text{ m}^{-1}$$

which comes very closely to the gradient of the excess refractive index for the standard troposphere, defined as

$$\frac{dN}{dh} = -4.3 \times 10^{-2} \text{ m}^{-1}$$

For practical purposes, it is often assumed that

$$\frac{dN}{dh} \cong -4 \times 10^{-2} \text{ m}^{-1}$$

From (3.7) it follows that N tends to zero owing to the monotonic decrease of pressure and humidity with height. An idealised relationship between the excess refractive index and altitude is shown in Fig. 3.1. This relation is usually referred to as the vertical profile of the excess refractive index. From the profile it follows that if the gradient of the excess refractive index were constant throughout

* This describes an ascending current of air when there is no heat exchange with the surroundings. The temperature drops as the air moves into a region of reduced pressure and reduced internal energy.

the atmosphere, the index would reduce to zero at an altitude of 8 kilometres. Actually, the lapse rate of N begins to decrease from an altitude of 7 kilometres, and a bend appears in the profile.

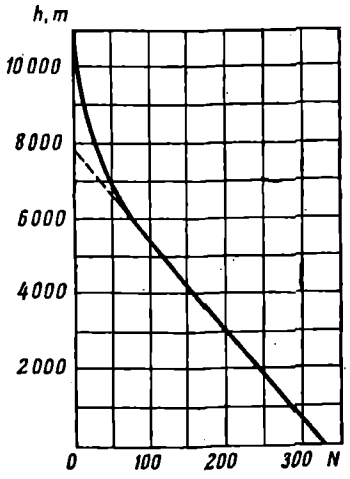


Fig. 3.1. An idealised refractive-index profile

Under actual conditions, there may be marked departures from the model profile, since the troposphere is greatly affected by the weather. Superimposed on the slow variations in the profile (due to changes in the weather) are the more rapid changes (or fluctuations) due to local irregularities in the refractive index. In cases of temperature inversions (discussed a bit later), N may sometimes decrease faster or more slowly than in the standard troposphere.

To sum up, the troposphere is an inhomogeneous dielectric whose refractive index varies with height, and so does, as a consequence, the velocity of wave propagation through it. The properties of this dielectric vary continually with time, so that the fluctuations related to the formation, movement and disappearance of local inhomogeneities in the troposphere are superimposed on the slower variations caused by meteorological conditions.

3.3. The Effect of the Troposphere on Ground Waves. Atmospheric Refraction

Inhomogeneities in the troposphere have a direct bearing on *atmospheric refraction* (first noted in the 2nd century B. C.). Atmospheric refraction consists in that a ray of light (and, by the same token, radio ray) encounters variations in atmospheric refractive index along its trajectory that cause the ray path to become curved.

Let us derive an expression for the radius of curvature of the ray path in the troposphere. For simplicity, we shall neglect the effect of the curvature of the earth's surface and assume that the troposphere consists of strata, or planes, each with a constant value of N and parallel to the flat earth. Consider two such planes (Fig. 3.2) spaced a distance dh apart. The refractive index of the lower plane or surface is n, and that of the upper surface, $n + dn$.

A ray, incident upon the lower surface at an angle φ and refracted along the distance dh, will fall upon the upper surface at an angle

$\varphi + d\varphi$. Since at point b the element of the ray path is turned through the angle $d\varphi$ relative to the element of the ray path at point a, the same angle is obtained between the normals to these path elements, that is, the angle at the centre 0 of the curvature.

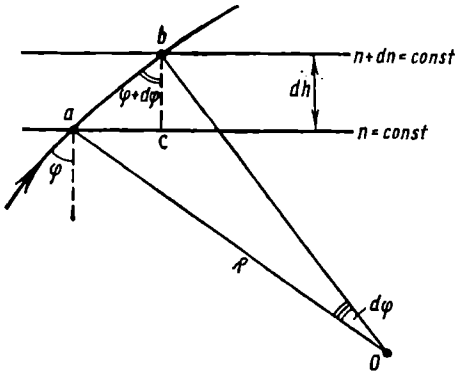


Fig. 3.2. Determining the radius of curvature of the wave path

The sought radius of curvature, R, is given by

$$R = \frac{ab}{d\varphi} \text{ m} \tag{3.8}$$

From the triangle abc we get

$$ab = \frac{dh}{\cos(\varphi + d\varphi)} = \frac{dh}{\cos\varphi} \text{ m}$$

whence,

$$R = \frac{dh}{\cos\varphi\, d\varphi} \text{ m} \tag{3.9}$$

In media with a gradually varying refractive index the law of refraction must hold at any point along the propagation path, including points a and b in our example. Therefore, we may write

$$n \sin\varphi = (n + dn) \sin(\varphi + d\varphi)$$

Writing out the right-hand side and neglecting the second-order infinitesimals, we get

$$n \sin\varphi = n \sin\varphi + n \cos\varphi\, d\varphi + \sin\varphi\, dn$$

or

$$\cos\varphi\, d\varphi = -\frac{\sin\varphi\, dn}{n}$$

Substituting it in (3.9) gives

$$R = \frac{n}{\sin\varphi \left(-\frac{dn}{dh}\right)} \text{ m} \tag{3.10}$$

Without any detriment to the accuracy of calculation, we may set $n \cong 1$ in the expression just derived. Furthermore, in connection with ground waves we are more often interested in rays propagated at small elevation angles for which $\sin \varphi \cong 1$ very nearly. As a result, the expression for the radius of curvature of the ray path may be further simplified as follows

$$R = \frac{1}{-dn/dh} = \frac{10^6}{dN/dh} \text{ m} \qquad (3.10a)$$

From (3.10a) it is seen that in the lower troposphere the radius of curvature of the ray path is decided by the lapse rate of refractive index with height, and not by its absolute value. The minus sign of the derivative implies that the radius of curvature will be positive, that is, the propagation path will be convex only when the refractive index decreases with height.

In the standard troposphere for which the gradient dN/dh is fixed, radio waves propagated at small elevation angles will travel in arcs of a circle whose radius is given by

$$R = \frac{10^6}{4 \times 10^{-2}} = 2.5 \times 10^7 \text{ m} = 25{,}000 \text{ km}$$

It should be noted that ultra-short waves undergo a greater refraction in the standard troposphere than light rays of optical frequency radio waves. This is because water molecules having a permanent dipole moment and a finite mass fail to follow the alternations of the electromagnetic field which occur at a very high rate in the visible region of the spectrum (4×10^{14} to 7.5×10^{14} hertz). In contrast, in the USW band, that is, at frequencies below 3×10^{11} hertz, polar molecules fully participate in the oscillatory motion and contribute to changes in refractivity. For optical-frequency radio waves and light rays, R is approximately 50,000 kilometres.

The refraction occurring in the standard troposphere will be referred to as *standard refraction*. Where doubt may arise, it is important to specify whether standard radio refraction or standard optical refraction is meant.

3.4. Effective Earth's Radius

The classical method of accounting for the atmospheric refraction of ground waves is to assume an effective earth's radius.

Let us consider this method, assuming that the transmitting and receiving aerials are elevated and that the propagation conditions are such that we may use the interference equations.

As will be recalled, this happens only in the USW band. The interference equations were derived on the assumption that both

the direct and reflected rays are propagated in straight lines at constant velocity. In reality, none of these conditions hold. Owing to atmospheric refraction, the paths of the direct and reflected rays are curved toward the earth, as shown in Fig. 3.3, which also shows the ray paths (the broken lines) without refraction. Therefore, the geometric path length difference in the interference equations will

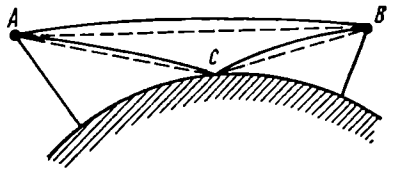


Fig. 3.3. A direct and a ground reflected ray in a real atmosphere (the full lines). The broken lines show straight-line trajectories in a homogeneous atmosphere

not be the same for straight and curved ray paths, and the attenuation function will likewise have different values. Besides, the propagation velocity in the lower troposphere (having larger values of n) will be lower than it is in the upper troposphere. Under the circumstances,

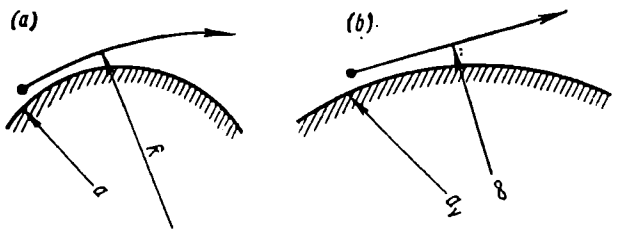


Fig. 3.4. Ray path: (*a*) over the true earth; (*b*) over the equivalent earth

the value of the attenuation function will be governed by the *optical path length difference* which in addition to the geometric path length difference allows for the difference in propagation velocity between the individual segments of the ray path.

In 1933, Schelling, Burrows and Ferrel [37] advanced a simple method which assumes an earth appropriately larger than the actual earth so that the curvature of radio ray may be absorbed in the curvature of the effective earth, thus leaving the relative curvature of the two the same and allowing radio rays to be drawn as straight lines over this effective earth rather than as curved lines over the true earth (Fig. 3.4).

In analytical geometry, the relative curvature is defined as the difference $1/a - 1/R$. Equating the relative curvatures for the situations shown at a and b in Fig. 3.4, we get

$$\frac{1}{a} - \frac{1}{R} = \frac{1}{a'} - \frac{1}{\infty}$$

whence the effective or equivalent radius of the earth is

$$a' = \frac{a}{1 - a/R} \text{ m} \tag{3.11}$$

Substituting (3.11) in (3.10a) gives

$$a' = \frac{a}{1 + a\frac{dN}{dh}10^{-6}} \text{ m} \tag{3.11a}$$

Writing the ratio of the effective to the true earth's radius as k (known as the *effective earth's radius factor*), we may write

$$k = \frac{a'}{a} = \frac{1}{1 + a\frac{dN}{dR}10^{-6}}. \tag{3.12}$$

For standard atmospheric refraction, $dN/dh = -4 \times 10^{-2}$ metres^{-1}. Substituting it, and also the numerical value of the true earth's radius, $a = 6.37 \times 10^6$, in (3.11a) and (3.12) gives

$$a' \cong 8{,}500 \text{ km}$$

$$k = 4/3$$

With the concept of the effective earth's radius, all equations derived in § 2.7 can be extended to an inhomogeneous atmosphere by replacing a with a'.

Above all, this is true of the equation describing the line-of-sight range. With allowance for atmospheric refraction, Eq. (2.70) takes the form

$$r_0 = \sqrt{2ak}\,(\sqrt{h_1} + \sqrt{h_2}) \text{ m} \tag{3.13}$$

Substituting the value of k for normal atmospheric refraction yields

$$r_0 = 4.12\,(\sqrt{h_{1m}} + \sqrt{h_{2m}}) \text{ km} \tag{3.13a}$$

that is, instead of 3.57 as in (2.70a), Eq. (3.13a) contains the factor 4.12.

It may be noted that for the optical waveband where the diffraction is less noticeable, the line-of-sight range is given by

$$r_0 = 3.83\,(\sqrt{h_{1m}} + \sqrt{h_{2m}}) \text{ km} \tag{3.13b}$$

In using the interference equations which do not contain the earth's radius explicitly, tropospheric refraction is accounted for

simultaneously with the curvature of the earth's surface. As will be recalled, the curvature of the earth's surface is allowed for by replacing the actual aerial heights with the reduced ones. In determining the reduced aerial heights analytically, the following equations may be used instead of (2.72a):

$$\left.\begin{aligned} h_1' &= h_{1(m)} - \frac{r_{1(\mathrm{km})}^2}{k12.8} \text{ m} \\ h_2' &= h_{2(m)} - \frac{r_{2(\mathrm{km})}^2}{12.8k} \text{ m} \end{aligned}\right\} \tag{3.14}$$

If the product of the reduced aerial heights and the grazing angle are taken from the plots of Figs. 2.42 and 2.43, the parameter p should be found by the following formula

$$p = \frac{r}{\sqrt{2a'h_1}} = \frac{r}{\sqrt{2kah_1}} \tag{3.15}$$

According to Vvedensky and Ponomarev [39], the effective earth's radius model may be used in conjunction with the interference equations for small elevation angles. Fock [40] has investigated atmospheric refraction in connection with diffractive propagation. According to him, in this case, too, atmospheric refraction may be accounted for, as in the line-of-sight propagation, by assuming an earth appropriately larger than the actual earth.

Thus, the effective earth's radius model applies to ground waves in all wavelength bands. Notably, the plots of Figs. 2.46 and 2.47 in § 2.8 have been constructed on the assumption that radio waves are propagated in vacuum around an earth of radius $\frac{4}{3}a$.

Example 3.1. Determine the effect of atmospheric refraction with reference to Example 2.3, for five values of the vertical gradient of refractive index: zero (no refraction), -0.02 m^{-1}, -0.04 m^{-1} (standard refraction), -0.08 m^{-1}, and -0.12 m^{-1}.

Solution: Substituting the numerical values of excess refractive index in Eqs. (3.10a), (3.11a), and (3.12) gives the radius of curvature R of the ray path, the effective earth's radius a', and the factor k. The respective results are tabulated in Table 3.1.

The distances to the reflection point as found by (2.75), are

$$r_1 = 7.14 \text{ km}$$
$$r_2 = 2.86 \text{ km}$$

The reduced aerial heights are found by Eq. (3.14). The results are summarized in Table 3.1.

Table 3.1. FACTOR k AND ATTENUATION FUNCTION VERSUS VERTICAL GRADIENT OF REFRACTIVE INDEX

$\frac{dN}{dh}$, m^{-1}	0	−0.02	−0.04	−0.08	−0.12
R, m	∞	5×10^7	2.5×10^7	1.25×10^7	8.33×10^6
a', m	6.37×10^6	7.50×10^6	8.54×10^6	1.3×10^6	2.71×10^7
k	1.0	1.18	1.34	2.04	4.26
h_1', m	21.0	21.6	22.0	23.0	24.1
h_2', m	9.4	9.5	9.6	9.7	9.9
F	0.25	0.26	0.27	0.28	0.30
F_{rms}, mV/m	7.4	7.7	7.85	8.4	8.95

The attenuation function F and the field strength E_{rms} at the receiver are found by (2.22) and (2.23), and in Table 3.1.

3.5. The Effect of Fluctuations in the Troposphere on Ground Wave Propagation

So far we have dealt with a time-invariant troposphere. Actually the state of the troposphere is varying continually, affecting the conditions of propagation and, as a consequence, the received field (see Example 3.1). An example is offered by the slow changes in the received field owing to variations in weather conditions.

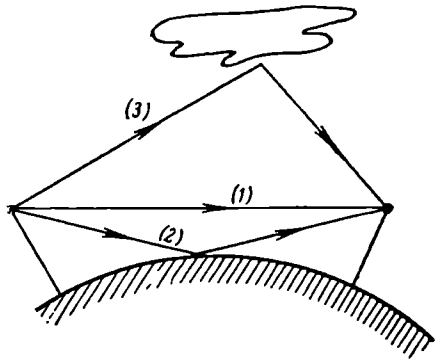


Fig. 3.5. Explaining the origin of fading

Micropulsations and fluctuations in the refractive index of the troposphere further complicate radio propagation. As stated more than once, the field within the line of sight is due to the combination of the direct and ground-reflected rays. The local irregularities

that these rays encounter along their trajectories, affect the velocity of propagation and, as a consequence, the resultant phase difference between the two rays. Phase fluctuations lead to fluctuations in the resultant field, known as *fading*.

In some cases, fading may be due to two direct rays (*1* and *2* in Fig. 3.5) interfering with a third ray, *3*, produced by reflection or scattering from an irregularity in the troposphere. The position of this irregularity varies rapidly, and so do the phase difference between the rays and the total received field.

A similar mechanism gives rise to fading over longer propagation paths where the ray reflected from an irregularity interferes with the diffracted field of the wave.

Since the ground-wave field is usually much stronger than that set up by a wave reflected or scattered from an irregularity, fading on short propagation paths is ordinarily shallow and introduces no significant signal distortion. However, fading becomes more of a problem on long microwave links owing to the cumulative effect of the individual hops.

Measures for fading control are discussed elsewhere in the book.

3.6. Forms of Atmospheric Refraction

In discussing the effect of tropospheric refraction on ground-wave propagation we had in mind an average condition referred to as the standard troposphere. However, weather conditions in the troposphere may lead to a refractive index distribution with height materially differing from this average state. Obviously, the effect of refraction will likewise be different.

In this section we shall consider several forms of refraction in the troposphere. Our classification will be based on the equations for N, dN/dh, R, a', and k. For convenience, three of these are written again:

$$R = \frac{10^6}{-dN/dh} \text{ m} \tag{3.10a}$$

$$a' = \frac{a}{1 - a/R} = \frac{a}{1 + a\dfrac{dN}{dh}\,10^{-6}} \text{ m}$$

$$k = a'/a \tag{3.12}$$

Above all, atmospheric refraction may be negative, zero, or positive. For each of these three forms Table 3.2 gives either numerical values of dN/dh, R, a', and k, or their limits. The last but one column presents sketches of ray paths for the various forms of refraction, and the last column illustrates propagation over the effective earth (that is over the 4/3 earth). In the latter case, the ray paths are straight-line segments.

Refraction is referred to as negative, when $dN/dh > 0$, that is, when instead of decreasing, N increases with height. As follows from (3.10a), in this case $R < 0$, that is, the ray paths are curved upwards. In other words, the radio wave moves away from the earth's surface, and the line-of-sight range and the range of propagation decrease accordingly. Negative refraction is an infrequent occurrence, but it has to be reckoned with.

Sometimes weather conditions may be such that N will remain fixed in an interval of altitudes. This will be zero refraction, dealt with in the second line of Table 3.2.

The subsequent lines in Table 3.2 describe several cases of positive refraction which occurs when N decreases with height, that is, when $dN/dh < 0$. The ray paths are now curved toward the earth and there is an increase in the range of propagation.

Positive refraction may in turn be divided into *sub-refraction* (when the bending of rays is smaller than in standard refraction), standard (already discussed and described in line C2 of the table), *augmented* (in which the bending of rays is greater than in standard refraction, but stops short of the critical one), *critical* (in which the radius of curvature of the ray path is equal to the earth's radius), and *super-refraction* (in which the bending of rays is more pronounced than in critical refraction).

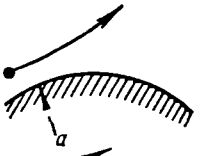
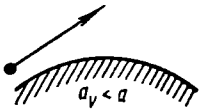
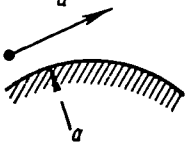
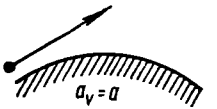
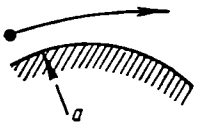
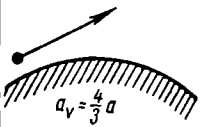
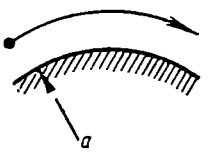
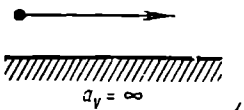
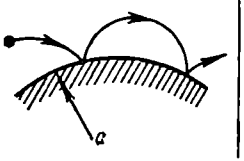
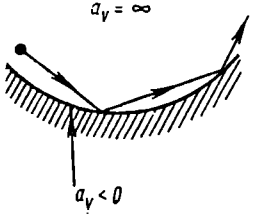
For a better understanding of this classification, we shall discuss critical refraction in greater detail. By definition, with critical refraction, $R = a$. Substituting the numerical value of the earth's radius for R in (3.10a) gives

$$\frac{dN}{dh} = \frac{10^6}{6.37 \times 10^6} = -0.157 \text{ m}^{-1} \text{ (line } C4 \text{ in Table 3.2)}$$

From Eq. (3.11) it follows that the effective earth's radius takes infinitely large values, that is, the earth reduces to a plane. Under conditions of critical refraction, a horizontally directed ray will be propagated at a fixed height above the earth's surface. That is, it will travel all the way around it.

With super-refraction, the radius of curvature of the ray path is smaller than the earth's radius, and the rays leaving the transmitting aerial at small angles of elevation will undergo total internal reflection in the troposphere and return to the earth at some distance from the transmitter (see the actual ray paths in Table 3.2). On reaching the earth's surface and being reflected from it, the waves can skip large distances, thereby giving abnormally large ranges beyond the line of sight due to multiple reflections. It is interesting to consider the respective ray path over the 4/3 earth. According to (3.11), when $R < a$, the effective earth's radius is a negative quantity. That is, the earth must be approximated by

Table 3.2. VARIOUS FORMS OF ATMOSPHERIC REFRACTION

No.	Form of refraction	$\frac{dN}{dh}$, m^{-1}	R, m	a_{eq}, m	k	Actual path	Equivalent path
A	Negative	>0	<0	$<6.37\times10^6$	<1		
B	Zero	0	∞	6.37×10^6	1		
C	Positive						
*C*1	Sub-refraction	0 to 0.04	∞ to 2.5×10^7	6.37×10^6 to 8.5×10^6	1 to 4/3		
*C*2	Standard	-0.04	2.5×10^7	8.5×10^6	4/3		
*C*3	Augmented	-0.04 to -0.157	2.5×10^7 to 6.37×10^6	8.5×10^6 to ∞	4/3 to ∞		
*C*4	Critical	-0.157	6.37×10^6	∞	∞		
*C*5	Super-refraction	<-0.157	$<6.37\times10^6$	<0	<0		

a concave surface. As can be seen from reference to the last heading of Table 3.2 for line *C*5, the straight rays, after multiple reflections from a concave surface, may reach extremely remote points.

It is not at all necessary that a particular form of refraction should prevail across the entire body of the troposphere. On the whole, the conditions in the troposphere are very close to standard, and the departures from the normal state, listed in Table 3.2, will only occur within particular intervals of height.

The refractive index of the air may be measured directly by a refractometer or it may be derived from measurements of the temperature, pressure, and humidity of the air by instruments carried aloft by radiosondes or aircraft. The values of T, p, and e thus found are then substituted in Eq. (3.7) to find N, and the excess refractive index is finally plotted as a function of height to produce a "reflective index profile".

Of many types of refractometer, the one most commonly used is a radio refractometer. One type consists essentially of two oscillators (usually of the klystron type) tuned to the same frequency in the SHF (centimetric) band. The cavity resonator of one oscillator (the reference cavity) is reliably sealed off from the atmosphere, while that of the other (the sampling cavity) is left open to the surrounding air. Therefore, the frequency generated by the other oscillator is affected by the refractive index of the air that fills it. The value of refractive index is determined from the difference in frequency between the two oscillators. The radio refractometer is usually carried aloft by aircraft, and its readings are used to obtain continuous records.

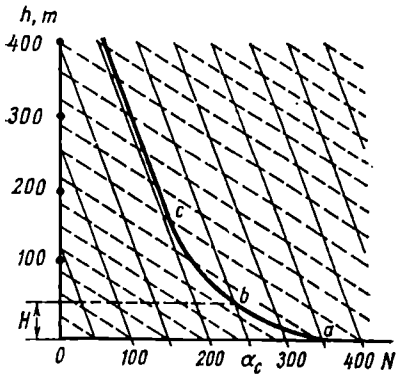


Fig. 3.6. An $N = f(h)$ profile. The full lines represent standard refraction, and the broken lines, critical refraction

Records of the excess refractive index are usually plotted on paper charts having two sets of parallel lines. One set (the full lines) represents standard atmospheric refraction ($dN/dh = -0.04$ m^{-1}), and the other (the broken lines) corresponds to critical refraction ($dN/dh = -0.157$ m^{-1}) (Fig. 3.6). Once the experimental $N = f(h)$ curve is plotted on this chart, it is compared with the respective parallel lines, and the extent of refraction is readily identified along with the interval of height where it occurs. In Fig. 3.6, the curve between *ab* has a greater slope than the broken lines. This

is an indication of super-refraction in that height interval. At point *b*, where a line tangent to the observed curve is parallel with the broken lines, critical refraction takes place. The portion *bc* corresponds to augmented refraction. Finally, at *c*, standard refraction occurs.

3.7. Conditions for Super-refraction

From the foregoing it follows that super-refraction may occur in a limited height interval in the troposphere where

$$\frac{dN}{dh} < -0.157 \text{ m}^{-1}$$

With super-refraction, the refractive index decreases at a rate which is four times the lapse rate of standard refraction.

Consider the weather conditions that are conducive to super-refraction. To begin with, let us differentiate (3.7) for N with respect to height by the rule of differentiation for complex functions

$$\frac{dN}{dh} = 77.6 \left[\frac{1}{T} \frac{dp}{dh} - \left(\frac{p}{T^2} + \frac{9620e}{T^3} \right) \frac{dT}{dh} + \frac{4810}{T^2} \frac{de}{dh} \right] \text{m}^{-1} \quad (3.16)$$

From Eq. (3.16) it follows that the vertical gradient of refractive index, dN/dh, at a height h and the lapse rate of this gradient are mainly determined by the gradients of pressure, temperature, and humidity. Although p, T, and h vary with height, their numerical values have a much smaller effect on the gradient of refractive index.

The pressure of the air decreases with height always, and its gradient is dependent on weather conditions only slightly. Therefore, the first term in (3.16) is nearly constant and always negative. In contrast, the gradients of temperature and humidity are subject to strong variations—they are markedly dependent on weather conditions and may even change sign (owing to so-called temperature inversions, inside clouds, and in "moisture pockets"). As already noted, in the standard troposphere the temperature and humidity always decrease with height, for which reason the two derivatives dT/dh and de/dh are negative. Therefore, the absolute value of dN/dh, which always takes a minus sign, is obtained by subtracting the absolute value of the second term from the sum of the absolute values of the first and third terms in the $(-) - (-) + (-)$ sequence.

Under certain weather conditions, what are known as temperature inversions may occur — instead of decreasing with height as is usual, the temperature increases within a particular interval. Obviously, with temperature inversion, $dT/dh > 0$, and the second term in (3.16) takes a minus sign. Now the absolute value of dN/dh is obtained by adding together three negative terms in the $(-) - (+) + (-)$ sequence.

Thus, among the conditions conducive to super-refraction, that is, abnormally high negative values of dN/dh, are above all temperature inversions and an extremely high lapse rate of humidity with height. Of the two factors, temperature inversion is the decisive one. There may be surface inversions and elevated inversions. The dependence of refractive index on height for surface inversion is plotted in Fig. 3.6, and for elevated inversion, in Fig. 3.7.

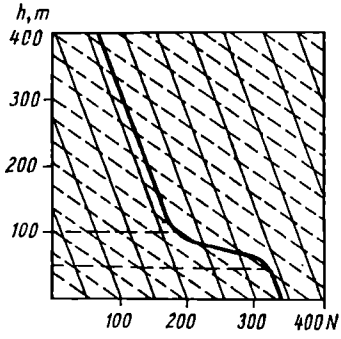


Fig. 3.7. Refractive index as a function of height in conjunction with elevated temperature inversion

Temperature inversions may be caused by advection processes, cooling of the earth's surface through radiation, and compression of air masses.

In an advection process, a body of warm dry air may be blown over a cooler layer of air (elevated inversion) or over a cooler underlying ground and, as a consequence, over a cooled layer of air (due to contact with the earth's surface). This type of inversion is to be observed in early spring, when masses of warm air are blown from the warmer regions of the earth over the ground still covered by snow.

Another example of advection is diagrammatically shown in Fig. 3.8. In the day-time, land is heated faster than sea owing to

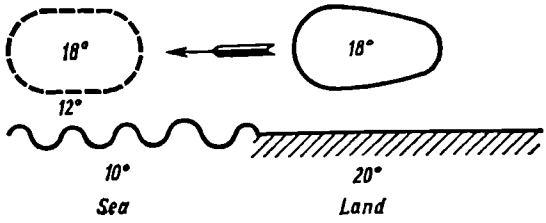


Fig. 3.8. Temperature inversion due to advection

the difference in specific heat (the numerals in Fig. 3.8 give arbitrary temperature). When a body of warm dry air blows from the land over the cooler sea, the air adjacent to the water grows colder, and a temperature inversion occurs. Besides, evaporation may take place or drops of water may be scooped from the crests of the waves, so that the humidity increases near the surface and decreases with height.

3.8. Radio Propagation under Conditions of Super-refraction

The ray-tracing approach to radio propagation under conditions of super-refraction (such as shown in Fig. 3.9) may only give a general qualitative picture of the process.

It is assumed that super-refraction affects a volume of the troposphere extending from the earth's surface out to an altitude h_0. The transmitting aerial is set up at point A. Rays starting at large elevation angles (*1* and *2* in the figure) undergo only partial refrac-

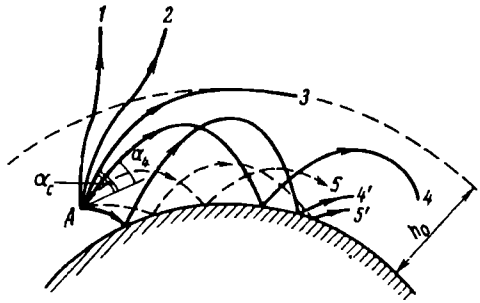


Fig. 3.9. Radio propagation in the case of super-refraction

tion and pass through the upper boundary of the super-refractive volume without being trapped there. Rays starting at an elevation angle α_{crit} become horizontal to the earth at an altitude h_0 where $dN/dn = -0.157$ metre^{-1} and where, according to the foregoing, the radius of curvature of the ray path is equal to the earth's radius. One such ray is labelled *3* in Fig. 3.9. All rays emitted within the sector $\pm\alpha_{crit}$ will be trapped by the super-refractive region, and abnormally large ranges beyond the line of sight are then obtained. One such ray, starting at an elevation angle α_4, is labelled *4* in the figure. This ray experiences total internal reflection at an altitude smaller than h_0, is reflected from the earth's surface, traces out the same path as before, and so on, the whole way along the super-refractive region in a horizontal direction. Obviously, it is possible to find a negative elevation angle $-\alpha_4'$ at which the ray will at first be reflected from the earth's surface and then trace out a path similar to the one just described. This path is marked *4'*. The broken lines represent rays *5* and *5'* starting at still smaller initial elevation angles and undergoing total internal reflection at lower altitudes.

Radio propagation under conditions of super-refraction resembles that in dielectric waveguides. The imperfectly conducting surface

of the earth acts as the bottom wall of such a waveguide, or *duct*, and the upper boundary of the super-refractive region, as its top wall. As with a dielectric waveguide, the refractive index inside the super-refractive region has a greater value than above it. The difference is that while in a dielectric waveguide the individual rays experience total internal reflection from both the top and bottom walls, in a super-refractive region they undergo ordinary reflection from the imperfectly conducting surface of the earth and total internal reflection from the super-refractive volume whose altitude is different for different angles of elevation.

Owing to this analogy, super-refractive regions in the troposphere are often referred to as *tropospheric waveguides* or *tropospheric ducts*, and radio propagation under conditions of super-refraction is called *ducting* or a *ducted mode*. The radio waves propagated over abnormally large ranges owing to ducting belong to *tropospheric waves*.

There is a still deeper analogy between dielectric and tropospheric waveguides. As with dielectric waveguides, the condition for the waves to be supported in a tropospheric duct is that the wavelength should not exceed λ_c, the cut-off wavelength. For the most commonly encountered cases, the following relationship exists between the cut-off wavelength and the height of the surface tropospheric duct

$$\lambda_c \cong 8.5 h_{0(\mathrm{m})}^{3/2} \times 10^{-4} \text{ m} \qquad (3.17)$$

The cut-off wavelengths satisfying (3.17) are given in Table 3.3.

Table 3.3. CUT-OFF WAVELENGTHS OF TROPOSPHERIC DUCTS AS FUNCTIONS OF DUCT HEIGHT

h_0, m	6	24	120	600
λ_c, m	0.01	0.1	1	10

From analysis of refractive index profiles based on observations it follows that the height of tropospheric ducts is usually measured in metres and tens of metres and almost never exceeds 200 metres. Thus, tropospheric ducts can support only ultra-short waves, mainly in the decimetric and centimetric bands. In rare cases, a tropospheric duct may guide waves two or three metres long. Tropospheric ducting may also have some effect on the propagation of metric waves exceeding the cut-off wavelength, thereby building up the received field. Ducting has no effect whatsoever on short, medium and long waves.

The ray-tracing approach to the tropospheric ducting of ultra-short waves gives only an approximate value of cut-off wavelength, but is useless in finding the attenuation function. Therefore, the attenuation function has to be found by solving Maxwell's equations for suitably specified boundary conditions.

A solution to the problem of radio propagation was obtained in 1946 by Krasnushkin [41] who used the normal-mode method. He postulated the possibility of ducting above the earth's surface due to weather effects before tropospheric ducts were observed experimentally.

This problem has been treated by several foreign authors including Booker and Walkinshaw [42], Pekeris and Ament [43].

In 1949, Brekhovskikh [44, 45] came out with a theory of guided wave propagation in stratified media, where the source is assumed to emit sets of plane waves which combine upon reflection from the individual strata.

As a technique for overland long-distance radio communications in the decimetric and centimetric bands, ducting is of limited value, because it is an extremely random occurrence. Conditions are more uniform, however, in some areas, such as the Arabian Sea and Trade Winds regions, and propagation by ducting can occur at certain hours of the day on a fairly regular basis.

As to super-refraction as such, it may result in unexpectedly abnormal ranges of propagation, introducing unforeseen interference. This is why it is important to allow for super-refraction in the siting of and frequency allocation for UHF and SHF transmitters.

3.9. Long-distance Propagation of Ultra-short Waves by Tropospheric Scattering

The diffraction equations examined in § 2.8 show that the absolute value of the attenuation function rapidly decreases with distance as the waves become shorter. This is why the range of propagation by diffraction is only slightly greater than the line-of-sight distance.

Figure 3.10 shows the attenuation function plotted against distance for four transmitted frequencies (30 megahertz, 300 megahertz, 3 gigahertz, and 30 gigahertz). As is seen, a wavelength of one centimetre ($f = 30$ gigahertz) is not practically diffracted around the earth. The curves of Fig. 3.10 have been plotted on the assumption that the transmitting and receiving aerials have the same height, $h_1 = h_2 = 10$ metres.

Numerous experiments have shown that the fields observed within 100 to 1000 kilometres of the transmitters are much greater than they ought to be according to the diffraction theory. In the same Fig. 3.10, the full lines represent the observed attenuation function

for 100 megahertz, 400 megahertz, 1 gigahertz, and 4 gigahertz. As is seen, the attenuation function does not drop below —100 decibels

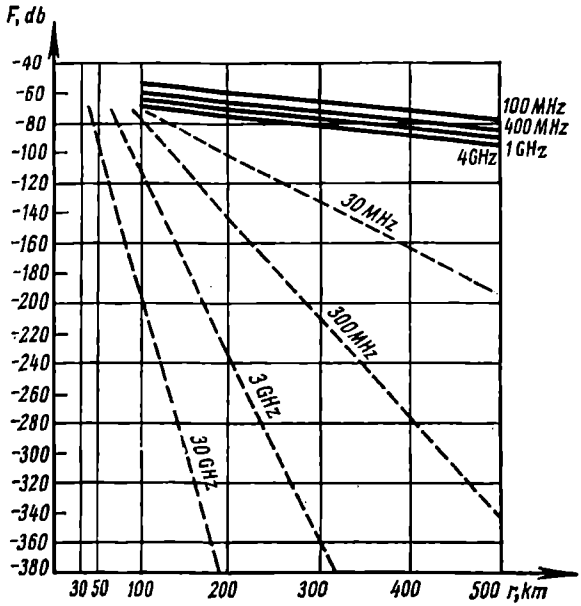


Fig. 3.10. Attenuation function plotted against distance. The full lines represent the observed values, and the broken lines, those calculated by the diffraction equations

even in the worst case at a distance of 500 kilometres, while even at 30 megahertz it is —190 decibels for radio propagation by diffrac-

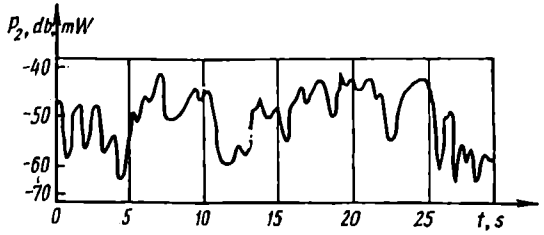


Fig. 3.11. Record of a signal subject to fading

tion. A distinction of the fields observed over long distances is that they are subject to fading, that is, random variations in strength caused by changes in the transmission medium.

Figure 3.11 shows an example of a signal record at 3.67 gigahertz, subject to fading. Time, in seconds, is laid off as abscissa, and the instantaneous power of the received signal, in decibels relative to one milliwatt, as ordinate. As is seen, the duration of fades varied from a split second to three seconds, and the depth of fading, defined as the ratio of the maximum signal level to the minimum one, was 40 decibels over 30 seconds. In spite of fading, the average signal level was sufficiently stable.

Fading is an indication that the fields set up over large distances are due not to ground waves, but to some different mechanism related to processes in the troposphere. It is natural to attribute fading to fluctuations in the troposphere.

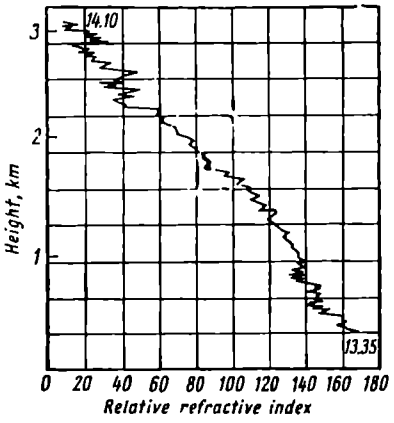


Fig. 3.12. A refractive-index profile plotted by a refractometer and showing its microstructure

Many facts, such as the twinkling of the stars, the wavering appearance of objects seen over the earth's surface heated by the sun, the shafts of smoke leaving tall stacks, and the conversion trails left by the exhaust gases of aircraft engines in the sky, all indicate that the air in the troposphere is in a random, erratic flow. The flow in which the local velocity of the air varies erratically with time both in direction and magnitude is called *turbulent*. The effect of turbulence on the variations in the refractive index with height can clearly be seen in profiles automatically plotted by radio refractometers. An example of such a profile is shown in Fig. 3.12. The serrations on the N (h) curves represent local irregularities in the refractive index of the troposphere caused by air turbulence.

The formation of turbulences, or eddies, is shown diagrammatically in Fig. 3.13. The reciprocal of eddy size, known as the mechanical wave number, $K = 2\pi/L = 2\pi/\Lambda$, is laid off as abscissa. The spectral density at the respective wavelength $\Lambda = L$ is laid off as ordinate. The spectrum may be divided into three ranges: (a) $0 < K < K_0$; (b) $K_0 < K < K_s$; and (c) $K > K_s$. In the range (a), as long as the local velocity of air remains below some critical value, the so-called laminar flow takes place — a particular type of streamline flow in which the air in thin parallel layers tends to maintain uniform velocity. At critical velocity, the laminar flow turns into a turbulent flow, large eddies of size L_0 are formed, and energy is fed into the system of eddies. In the range (b), the energy is conserved (because

of which this is called the *inertial region*), but the eddies decay in size gradually and continually until the lower size limit L_S is reached. In the range (c) the eddies are small and their energy is dissipated due to the viscosity of the medium, as shown by the arrows in Fig. 3.13. The spectrum of eddy sizes may be described as

$$L_0 > L > L_S \text{ m} \tag{3.18}$$

Owing to turbulent variations in pressure, temperature and humidity in the atmosphere there are also random variations in the permittivity. The value of the permittivity is a function of time

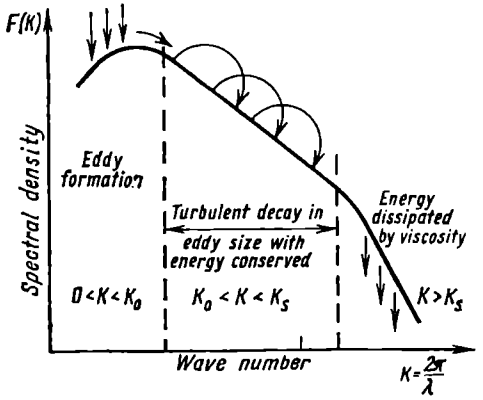


Fig. 3.13. Spectrum of refractive index variations

and the point of measurement. Therefore, the permittivity of a turbulent atmosphere may be written as

$$\varepsilon'(\mathbf{r}, t) = \varepsilon'_0 + \Delta\varepsilon'(\mathbf{r}, t) \tag{3.19}$$

where $\mathbf{r}$ is the radius vector from a reference point to the point of measurement, t is time, ε'_0 is the average permittivity, and $\Delta\varepsilon'$ is incremental permittivity (or turbulent variations in the average permittivity).

The value of ε' is usually regarded as being a stationary, zero mean, random function of space and time. It is assumed that the medium is isotropic and homogeneous, that is, the statistical properties of ε' are independent of position and direction. Under the circumstances, the statistical relationship between turbulent fluctuations in permittivity at points *1* and *2* spaced distance ρ apart is described by the spatial correlation function $C(\rho)$

$$C(\rho) = \frac{\overline{\Delta\varepsilon'_1(t)\,\Delta\varepsilon'_2(t)}}{\overline{(\Delta\varepsilon')^2}} = \lim_{T\to\infty} \frac{1}{2T\overline{(\Delta\varepsilon')^2}} \int_{-T}^{+T} \Delta\varepsilon'_1(t)\,\Delta\varepsilon'_2(t)\,dt \tag{3.20}$$

where the overscribed bar indicates a time average. The term $(\overline{\Delta\varepsilon'})^{\frac{1}{2}}$ is the mean intensity (or mean square) of the variations in amplitude of ε'.

By the ergodic theorem, a time average may be replaced with an *ensemble average*. That is, instead of correlating two events occurring at points *1* and *2* at different times, we may correlate events occurring at one and the same time at a multitude of points within the sample space, separated by the distance ρ. The correlation function then is

$$C(\rho) = \frac{1}{V\,\overline{(\Delta\varepsilon')^2}} \int_V \Delta\varepsilon'(\mathbf{r})\,\Delta\varepsilon'(\mathbf{r}+\rho)\,dV \qquad (3.21)$$

where **r** is the radius vector from a reference point to the point of measurement, varying within the sample space V.

The scale length, or average size, of the eddies in the turbulence is given by

$$l = \int_0^\infty C(\rho)\,d\rho \text{ m} \qquad (3.22)$$

As is seen from (3.22), the average size, or scale length, of a turbulence can be determined only if the correlation function is known. Physically, the quantity l determines, to a first approximation, the radius of a sphere within which the statistical relationship is strong. If points *1* and *2* are separated by a small distance, the variations at them will occur synchronously, and $C(0)$ will tend to unity. In the other limiting case, when the distance between the points is very great, the variations at one will be independent of those at the other, and $C(\infty)$ will tend to zero. Thus, $C(\rho)$ is a monotonically decreasing, even function of distance.

To sum up, the variables describing a turbulent troposphere are the mean square variation $(\overline{\Delta\varepsilon'})^2$ and the scale length l.

Much as an arbitrary time function meeting some specific conditions can be represented by a double Fourier integral, that is, as a continuous spectrum of harmonics, $\cos \omega t$, where ω is the radian frequency, so the monotonic function of distance ρ may be presented as a continuous spectrum of spatial harmonic components of the form $\cos 2\pi\rho/\Lambda$, where Λ is the length of a wave in space*.

Denoting the spatial distribution density (that is, the spatial or energy spectrum) as $F(K)$, where $K = 2\pi/\Lambda$ is the mechanical

* Λ is not to be confused with the length of a radio wave such that $\lambda = c/f$, where $c = 3 \times 10^8$ metres per second; it is the length of mechanical waves into which the eddies of air can be resolved.

wave number, we may write on the basis of Fourier's theorem

$$C(\rho) = \int_0^{\infty} F(K) \cos K\rho \, dK = \frac{1}{2} \int_{-\infty}^{+\infty} F(K) \, e^{iK\rho} \, dK \qquad (3.23)$$

Using the inverse Fourier transform gives

$$F(K) = \frac{2}{\pi} \int_0^{\infty} C(\rho) \cos K\rho \, d\rho = \frac{1}{\pi} \int_{-\infty}^{+\infty} C(\rho) \, e^{-iK\rho} \, d\rho \text{ m} \qquad (3.24)$$

Thus, scatter propagation in the troposphere may be equally treated in terms of both the correlation function and the energy spectrum. For typical turbulent states, the spatial spectrum has the form shown in Fig. 3.13. The size of the largest eddies is $L_0 = \Lambda_0 =$

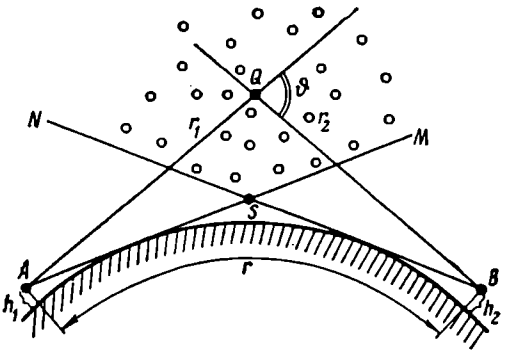


Fig. 3.14. Tropospheric-scatter radio link

$= 2\pi/K_0$ and that of the smallest eddies destroyed by viscous forces is $L_S = \Lambda_S = 2\pi/K_S$.

In the light of the foregoing it is not difficult to understand the mechanism by which ultra-short waves are propagated beyond the horizon due to tropospheric scattering. Fig. 3.14 shows diagrammatically a tropospheric-scatter radio link. The transmitting aerial of gain G_1 is set up at point A, at height h_1 above the earth's surface while the receiving aerial of gain G_2 is located at point B, at height h_2 above the earth's surface. The propagation path length is r. The straight lines AM and BN, drawn from the points of transmission and reception respectively, are tangent to the earth's surface. All points located above the tangent planes AM and BN have the property of being simultaneously seen from the terminals of the link. The collection of these points within the troposphere makes up what

is known as the *common* or *scattering volume*. All tropospheric irregularities within this common volume, when illuminated by the transmitting aerial, become sources of scattered radiation reaching the receiving aerial. The multitude of scattered-radiation sources produces the total field at the receiver. The combination of incoming waves gives rise to fading which always accompanies the reception of tropospheric-scatter signals.

In the Soviet Union the effect of refractive-index variations on the propagation of ultra-short waves was investigated for the first time by V. A. Krasilnikov [46] in 1949.

Consider a point, Q, within the common volume (Fig. 3.14) and let this point be r_1 distant from the transmitter and r_2 distant from the receiver. The straight lines AQ and QB make what is called the *scattering angle*, θ. It should be remembered that for clarity the vertical scale in Fig. 3.14 has been exaggerated. For a propagation path length, r, equal to 300 kilometres, the height of the lowest point S of the common volume will be one kilometre above the earth's surface, and the height of the uppermost point of the common volume, significantly contributing to the total scattered field at the receiver, will be about 10 kilometres. Therefore, in a first approximation, we may take it that all points of the common volume are within the main lobe of the radiation pattern of the aerial.

The power flux density at Q due to the transmitting aerial is given by

$$S_Q = \frac{P_1 G_1}{4\pi r_1^2} \text{ W/m}^2 \tag{3.25}$$

Let σ denote the effective scattering cross-section per unit volume in the troposphere, with Q as centre. As will be recalled, the effective scattering cross-section of a body is a quantity having the dimensions of area, which, when multiplied by the power flux density of the primary radiation (at the location of the scattering body) and divided by $4\pi r_2^2$ (where r_2 is the distance of the scattering body from the receiver), gives the scattered power flux density at the point of reception. Thus, the product $S_Q\sigma$ describes the re-radiated power of the body. The scattered field strength is a function of the scattering angle θ; therefore, it is convenient to denote the effective scattering cross-section as $\sigma(\theta)$. Besides, since we are dealing with the scattering cross-section per unit volume, the dimensions of the function $\sigma(\theta)$ are square metre $\div$ cubic metre = 1/metre.

By definition, the re-radiated power due to a volume element dV is given by

$$dP_Q = \frac{P_1 G_1 \sigma(\theta)}{4\pi r_1^2} dV \text{ W} \tag{3.26}$$

The power flux density of the wave at the receiver is

$$dS_B = \frac{dP_Q}{4\pi r_2^2} \text{ W/m}^2 \tag{3.27}$$

and the power extracted by the receiving aerial is

$$dP_2 = dS_B A_2 \text{ W} \tag{3.28}$$

where A_2 is the effective area of the receiving aerial which, on the strength of (1.17), may be expressed in terms of the aerial gain G_2 as

$$A_2 = \frac{G_2\lambda^2}{4\pi} \text{ m}^2 \tag{3.29}$$

Substituting the expressions for dP_Q and A_2 in (3.28) gives

$$dP_2 = \frac{P_1 G_1 G_2 \sigma(\theta)\lambda^2}{16\pi^2 r_1^2 r_2^2 4\pi} \text{ W} \tag{3.30}$$

To find the power due to the entire scattering volume, we must integrate (3.30) over that volume

$$P_2 = \frac{P_1 G_1 G_2 \lambda^2}{64\pi^3} \int \frac{\sigma(\theta)\,dV}{r_1^2 r_2^2} \text{ W} \tag{3.31}$$

If radio waves were propagated in free space, the power at the receiving aerial, according to (1.19) and with the aerials properly oriented, would be

$$P_{2f} = \frac{P_1 G_1 G_2 \lambda^2}{(4\pi r)^2} \text{ W} \tag{3.32}$$

Denoting the field attenuation function as F, the actual power at the receiving aerial may be written as

$$P_2 = P_{2f} F^2 = \frac{P_1 G_1 G_2 \lambda^2 F^2}{(4\pi r)^2} \text{ W} \tag{3.33}$$

From a comparison of (3.31) and (3.33), the attenuation function may be written as

$$F = \frac{r}{2\sqrt{\pi}} \sqrt{\int_V \frac{\sigma(\theta)}{r_1^2 r_2^2} dV} \tag{3.34}$$

When highly directive aerials are used and the common volume in the troposphere is at the intersection of the radiation patterns of both aerials, r_1 and r_2 will vary insignificantly within the scattering volume and may be placed outside the integral sign. Besides, we may approximately set

$$r_1 = r_2 \cong r/2 \tag{3.35}$$

Substituting this in (3.40) gives

$$F \cong \frac{2}{r\sqrt{\pi}} \sqrt{\int_V \sigma(\theta)\, dV} \tag{3.36}$$

Finally, if the scattering volume is so small that the function $\sigma(\theta)$ may be taken constant within its boundaries, Eq. (3.36) will reduce to

$$F \cong \frac{2}{r\sqrt{\pi}} \sqrt{\sigma(\theta)\, V} \tag{3.37}$$

From Eqs. (3.34), (3.36), and (3.37), it follows that for the attenuation function to be found analytically, we must know: (1) the structure of the scattering volume V in the troposphere, and (2) the expression for the effective scattering cross-section.

3.10. Determination of the Effective Scattering Cross-section and the Significant Scattering Volume in the Troposphere

Writing the radio wave number as $k = 2\pi/\lambda$, the mechanism of scattering by a unit volume of the troposphere may be pictured as shown in Fig. 3.15.

Here, $\mathbf{k}_1$ is the vector of the length k oriented in the direction of the incident wave, and $\mathbf{k}_2$ is the vector of the same length oriented in the direction of the scattered wave. As already noted, the angle between the vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ is the scattering angle.

For clarity, the vertical scale in Fig. 3.15 has been exaggerated. Actually, the angle θ is from a fraction of a degree to several degrees.

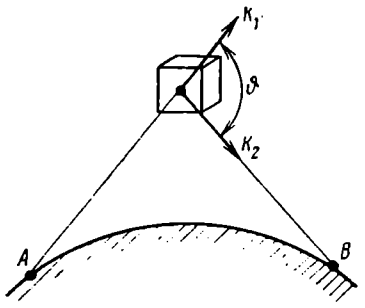


Fig. 3.15. Mechanism of scattering by a unit volume in the troposphere

Let us show that scattering in the direction of θ is determined by the vector difference $\mathbf{k}_1 - \mathbf{k}_2$. We denote the absolute value of the difference as K and refer to Fig. 3.16 which illustrates an idealized case of propagation where scattering elements are located in parallel planes *1-1'*, *2-2'*, *3-3'*, etc., spaced the eddy size L apart. Under these conditions, the ray scattered at point B will cover an additional propagation path $mB + Bn$ as compared with the ray scattered at point A. As a consequence, there will be an additional phase shift

$$\Delta\varphi = \frac{2\pi}{\lambda}(mB + Bn) = k(mB + Bn) \tag{3.38}$$

From Fig. 3.16, it follows that

$$mB = Bn = L \sin \frac{\theta}{2} \text{ m} \tag{3.39}$$

Substituting this in (3.38) gives

$$\Delta\varphi = 2kL \sin \frac{\theta}{2} \tag{3.40}$$

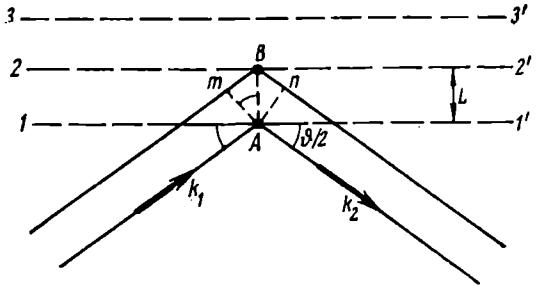


Fig. 3.16. Propagation with scattering elements located on parallel planes spaced the eddy size apart

On the other hand, from the vector diagram appearing in Fig. 3.17 it follows that

$$|\mathbf{K}| = |\mathbf{k}_1 - \mathbf{k}_2| = 2k \sin \frac{\theta}{2} \text{ (m}^{-1}\text{)} \tag{3.41}$$

Using this expression, we may re-write Eq. (3.40) as

$$\Delta\varphi = KL \tag{3.40a}$$

The most favourable conditions for wave scattering will be obtained when the radiation scattered from the elements in plane *1-1'*

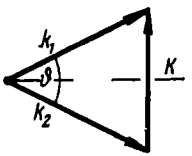


Fig. 3.17. Diagram of the vector difference $\mathbf{k}_1 - \mathbf{k}_2$

is in phase with that due to the elements in planes *2-2'*, *3-3'*, etc. Obviously, this will happen when

$$\Delta\varphi = KL = \frac{4\pi}{\lambda} \sin \frac{\theta}{2} L = 2\pi \tag{3.42}$$

or

$$2 \sin \frac{\theta}{2} L = \lambda \text{ m} \tag{3.42a}$$

Eq. (3.42) may be re-written as

$$K = 2\pi/L \text{ (m}^{-1}\text{)} \tag{3.42b}$$

Equations (3.42) and (3.42a) express the Bragg law, known from crystallography and describing the condition under which a crystal will diffract X-rays.

If the scattering angle θ and the wavelength λ are held constant, the effective scattering in the direction of the receiving aerial will

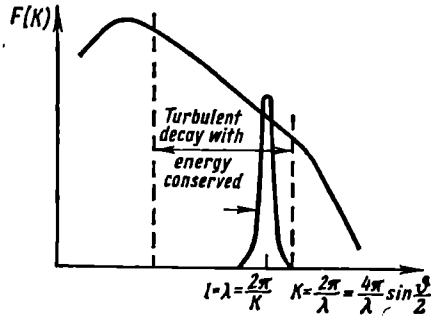


Fig. 3.18. Spatial spectrum of fluctuations and the pass-band of a narrow-band filter

be decided only by the elements the vertical separation between which is L as found from (3.42a). As already noted, the eddies present in a turbulence vary in size from L_0 to L_S. Therefore, for a given propagation path length (that is, for a given scattering angle θ) and a given wavelength the scattering will be determined by eddies of size L.

Because of this, we may liken scatter propagation to operation of a narrow-band filter which, with specified values of λ and θ, uses a particular size of irregularities, and these are mainly responsible for the scattered field at the point of reception. Fig. 3.18 shows the spatial spectrum shown earlier in Fig. 3.13 and also the pass-band of a narrow-band filter. The numerals along the X-axis indicate the eddy sizes mainly contributing to scattering.

The quantity L is in effect the spatial period of tropospheric irregularities or, which is the same, the wavelength of mechanical oscillations into which the turbulent flow may be resolved. Noting that this quantity was earlier designated as Λ, we may re-write (3.42b) as

$$K = 2\pi/\Lambda \text{ m}^{-1} \tag{3.42c}$$

Thus, the quantity K entering the Bragg equation for turbulent scattering has the same meaning as the mechanical wave number.

The statistical approach to tropospheric scattering of radio waves yields the following expression for the effective cross-section per unit of the scattering volume [47]

$$\sigma = \frac{\sin^2 \alpha \, \overline{(\Delta\varepsilon')^2} \, k^4}{4\pi} \int_0^\infty C(\rho) \, \rho^2 \, \frac{2 \sin K\rho \, d\rho}{K\rho} \ \text{m}^{-1} \qquad (3.43)$$

From (3.43) it follows that, once the correlation function $C(\rho)$ and the mean square variation in permittivity are known, we may uniquely determine the expression for the effective scattering cross-section σ.

Using (3.24), which relates the spatial fluctuation spectrum and the correlation function, we may give (3.43) a slightly different form. Thus, by the rule for differentiation under the integral sign, from (3.24) we get

$$\frac{1}{K} \frac{dF(K)}{dK} = -\frac{2}{\pi} \int_0^\infty C(\rho) \, \frac{\sin K\rho}{K\rho} \, \rho^2 \, d\rho \ \text{m}^2 \qquad (3.44)$$

and (3.43) takes the form

$$\sigma = -\frac{\sin^2 \alpha \, \overline{(\Delta\varepsilon')^2} \, k^4}{8} \frac{dF(K)}{dK} \cdot \frac{1}{K} \ \text{m}^{-1} \qquad (3.45)$$

That is, the scattering cross-section for tropospheric scattering may uniquely be expressed in terms of the mean square variation and the energy spectrum.

At present, a generally recognized theory of tropospheric scattering is still non-existent, and different authors give various different mathematical expressions for the correlation function or, which is the same, the energy spectrum. Because of this, (3.45) may take several specific forms.

In the region of the spectrum where $F(K) = AK^{-n}$, the term

$$\frac{dF(K)}{dK} \frac{1}{K}$$

in the expression for the scattering cross-section will be

$$-nAK^{-(n+2)} = -nA \left(2k \sin \frac{\theta}{2}\right)^{-(n+2)} \ \text{m}^3 \qquad (3.46)$$

Substituting (3.46) in (3.45) yields

$$\sigma = \frac{An\pi^2 \sin^2 \alpha}{8 (4\pi)^n} \overline{(\Delta\varepsilon')^2} \, \lambda^{(n-2)} \left(\sin \frac{\theta}{2}\right)^{-(n+2)} \ \text{m}^{-1} \qquad (3.47)$$

Eq. (3.47) is interesting in two respects. To begin with, it shows that scattering is strongest in the direction of the wave emitted by

the transmitter. In this respect, scattering from local irregularities in the troposphere markedly differs from that due to metallic bodies. As will be recalled, a metallized sphere will scatter the energy of the incident radio wave uniformly in practically all directions. This is why tropospheric scattering is often referred to as *forward*. As the scattering angle increases, the intensity of scattered radiation drastically decreases. Besides, Eq. (3.47) describes tropospheric scattering as a function of wavelength.

The mixing-in-gradient theory of turbulence due to Villars and Weiskopf [48], [49] assumes an initial gradient of permittivity with height. Although turbulent mixing causes pockets of air to be transformed from one level to another, the permittivity varies about a certain mean level. This results in $d\varepsilon'/dh$, the steady vertical gradient of permittivity, such that

$$\Delta\varepsilon' \cong \left(\frac{d\varepsilon'}{dh} - \overline{\frac{d\varepsilon'}{dh}}\right) L \tag{3.48}$$

where the term in brackets describes fluctuations in the vertical permittivity gradient of the troposphere.

Thus, instead of the term $\overline{(\Delta\varepsilon')^2}$ in Eq. (3.45) for the effective scattering cross-section, the mixing-in-gradient theory uses

$$\overline{(\Delta\varepsilon')^2} = \left(\overline{\frac{d\varepsilon'}{dh} - \overline{\frac{d\varepsilon'}{dh}}}\right)^2 L^2 \tag{3.48a}$$

where L is found from (3.42b) and (3.42c)

$$L = \Lambda = \frac{2\pi}{K} \frac{\lambda}{2\sin\frac{\theta}{2}} \text{ m} \tag{3.49}$$

Substituting this in (3.48a), and the latter in (3.47), we get

$$\sigma = B\sin^2\alpha \left(\overline{\frac{d\varepsilon'}{dh} - \overline{\frac{d\varepsilon'}{dh}}}\right)^2 \lambda^n \left(\sin\frac{\theta}{2}\right)^{-(n+4)} \tag{3.50}$$

where B is the proportionality factor.

A more recent formulation of the Villars-Weiskopf theory [49] assumes n to be equal to unity, and the quantity σ becomes

$$\sigma = B\sin^2\alpha \left(\overline{\frac{d\varepsilon'}{dh} - \overline{\frac{d\varepsilon'}{dh}}}\right)^2 \frac{\lambda}{\sin^5\frac{\theta}{2}} \text{ m}^{-1} \tag{3.51}$$

In this case, the effective scattering cross-section is proportional to the first power of the wavelength, which agrees well with experimental observations. In experiments, it has been observed that the mean square variation in the permittivity gradient with height may be plotted as shown in Fig. 3.19.

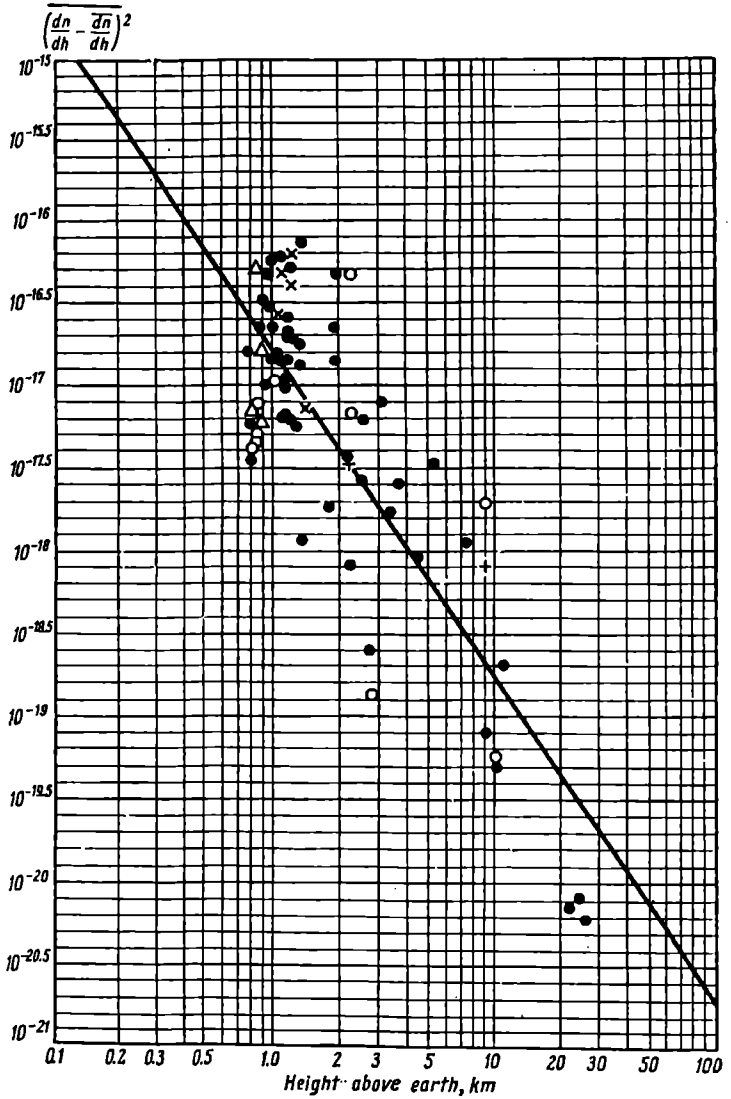


Fig. 3.19. Mean square variation in permittivity gradient with height

Other authors are of the opinion that the trans-horizon propagation of ultra-short waves in the troposphere is mainly due to partial reflection from layered irregularities at the sharp boundaries of clouds, at the interface between warm and cold air masses, and to some other meteorological factors.

Diagrammatically, the total field at the point of reception due to layered irregularities is shown in Fig. 3.20. The irregularities may

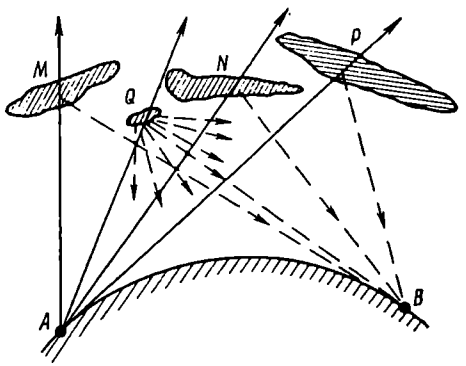


Fig. 3.20. Received field resulting from reflections and scattering by layered irregularities in the troposphere

widely differ in shape, size, and orientation. Those extending for a considerable distance give rise to specular reflections of radio waves (M, N, and P in Fig. 3.20), while small irregularities (Q) cause a scattering similar to that produced by local irregularities in the theory just discussed. The multitude of layered irregularities rapidly moving and changing in shape results in wave interference and, as a consequence, fading.

For the first time, a theory of scattering from layered irregularities was set forth by V. N. Troitsky in 1956 [50]. Later, it was expounded by Friis, Crawford, and Hogg [51].

As regards the determination of the scattering volume in the troposphere, two cases ought to be examined—aerials of low directivity and aerials of high directivity.

With aerials of low directivity, the scattering volume is decided by the fact that the effective scattering cross-section σ drastically decreases as the scattering angle θ increases, because the angle θ in the denominator is raised to fifth power. It is assumed that the sines may well be replaced with their arguments because the angles are very small. Besides, the mean square variation in permittivity (see Fig. 3.19) is drastically decreased with height.

In this case the common scattering volume is bounded from below by two tangent planes containing the transmitter and receiver and normal to the plane drawn through the terminal points of the propagation path and the centre of the earth and coincident with that of Fig. 3.14. A more detailed analysis would show that on the sides the common scattering volume is bounded by planes parallel to the plane of Fig. 3.14 and $5h_0$ distant from it, where h_0 is the height of the lowest point of the scattering volume (point S in Fig. 3.14). From above, the scattering volume is bounded by a plane horizontal to the point located midway along the propagation path and $4h_0$ above that point.

With highly directional aerials, the common scattering volume is produced by the intersection of the radiation patterns of the transmitting and receiving aerials.

3.11. Fading in Tropospheric Scatter Propagation

As already noted, ultra-short waves propagated by tropospheric scatter are subject to fading.

Fading proper applies to the continuous, rapid variations in radio field strength occurring at random at the point of reception, thereby producing changes in the signal level and introducing distortion. These variations may last from several minutes down to seconds or even a few fractions of a second. They ought not to be confused with random variations in the average signal level of longer duration on which fading proper is superimposed. In turn, these random variations ought not to be confused with regular diurnal and seasonal variations in the radio field strength observed under various propagation conditions.

Fading occurs not only in tropospheric propagation, but also in some other instances. This is why the material presented in this section applies to all forms of fading, including that in ionospheric propagation.

The physical basis of fading is *multi-pathing*, that is, the fact that the received field is set up as a result of interference between waves which have reached the receiver by different paths. This is why in tropospheric scatter propagation of ultra-short waves fading is a natural companion—local irregularities in the troposphere scatter radio rays in all directions so that they reach the receiver by different paths.

Random variations in the average signal level in tropospheric propagation are due to meteorological causes—changes in the mean square variation in permittivity, changes in the conditions of turbulent mixing, fronts, variations in the temperature distribution in the troposphere, and so on. It is important to note that these more

slow changes in the field strength have nothing to do with wave interference.

Because of fading, we have to introduce suitable definitions for the average signal level and the deviation of instantaneous signal level values from the average one. Most commonly, the average signal level is expressed in terms of the *median* field strength. The median signal level is defined as that which is exceeded during a half of the reception time. Suppose that a signal is received during a time T, and that the variations in the field strength with time can be

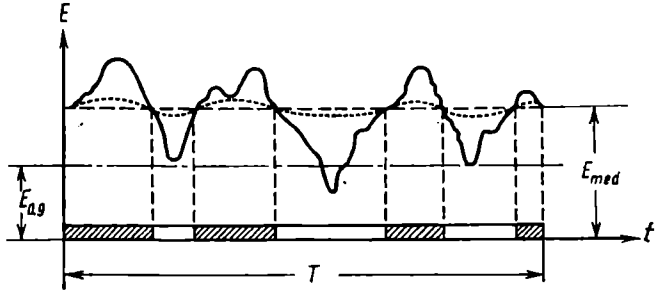


Fig. 3.21. Median signal level

represented by a full line as shown in Fig. 3.21. To find the median field strength, we draw a straight line parallel to, and distant from the X-axis, such that the sum of time intervals during which the actual field strength exceeds a specified level will be equal to the sum of time intervals during which the actual field strength is lower than the specified one. In Fig. 3.21, the median signal level is denoted by E_{med}. The time intervals during which this level is exceeded are shaded. The total length of the shaded segments is equal to the total length of the unshaded segments.

Although it gives an idea about the average level of the signal, the median field strength does not show how deep the fading is. Two signals with the same median value may be subject to fading differing in depth. The broken line in Fig. 3.21 represents a signal whose level intersects the median values at the same points as the signal discussed before (the full line). In other words, both signals have the same median value, but the one represented by the broken line does not practically fade out. For an evaluation of fading, it is important to specify additionally two or three levels which are exceeded during, say, 90, 99, and 99.9 per cent of the time of reception. One such level, namely $E_{0.9}$, is plotted in Fig. 3.21, and the respective median level is symbolized as $E_{0.9}$.

Fading may be described in terms of its average rate (the number of fade-outs per minute or per second) and the depth of fading.

The rate of fading is usually determined by counting the intersections of the median level. For example, in Fig. 3.21, the median level is crossed seven times over the time interval T; this works out to a rate of 3.5 cycles per time T.

The depth of fading can be evaluated in very rough terms only. To begin with, it cannot be defined as the ratio of the maximum to minimum field strength over the time interval T, because although some peaks may attain high values, the probability that they will

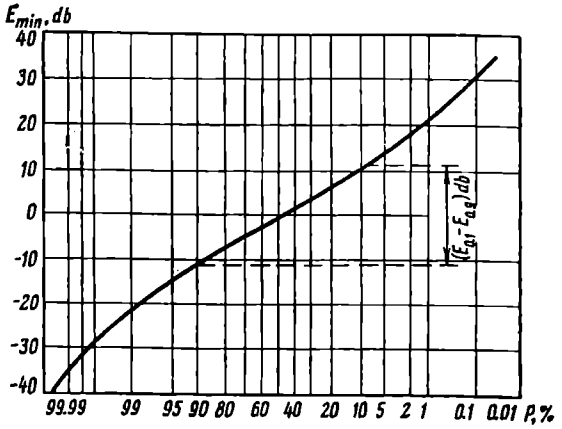


Fig. 3.22. Plot showing the distribution function of the received signal in the presence of fading

do so is very low. On the other hand, deep minima are usually masked by background noise. It is customary, therefore, to define the depth of fading as

$$(E_{0.1} - E_{0.9})$$

where $E_{0.1}$ is the field level exceeded during 10 per cent of the time and $E_{0.9}$ is the field level exceeded during 90 per cent of the time. In both cases, the field strength is expressed in decibels relative to one microvolt per metre, and the difference is also in decibels.

Fading is more completely described by the distribution function. An example of this function is shown in Fig. 3.22, where the time during which the signal exceeds a specified level is expressed as the percentage of the total reception time T, and is laid off as abscissa while the specified signal level is laid off as ordinate. The specified level may be taken in decibels relative to either one microvolt per metre, which is preferable, or to the median level arbitrarily assigned the value of zero decibels. A plot of the distribution function offers a simple tool for determining the depth of fading as the difference $(E_{0.1} - E_{0.9})$. The corresponding construction is shown by the

broken line in Fig. 3.22. The level distribution function offers exhaustive information about the nature of fading. Among other things, if we know the median level, in microvolts per metre, and the field level, also in microvolts per metre, which still secures good reception, we can determine the probable time of reliable reception, that is, the time during which the signal will exceed the threshold value.

Analysis of the slow changes in the signal propagated through tropospheric scatter indicates that the signal amplitude has the logarithmically normal probability distribution, while random variations in amplitude and phase, or fading, are described by the Rayleigh distribution.

As will be recalled, a large number of the random processes dealt with in science and engineering have a normal distribution. This distribution manifests itself in cases where the random variable of interest is subject to the influence of a large number of random variables each of which contributes only insignificantly to the result.

When a random variable Z has a normal distribution, its probability density is given by the function

$$p(Z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(Z-a)^2}{2\sigma^2}} \quad (\text{in } Z^{-1} \text{ units}) \tag{3.52}$$

where a = expectation or mean of the random variable Z (in Z units);

σ = standard deviation of Z (in Z units);

σ^2 = variance of Z, or the second-order moment of Z about the centre of the distribution (in Z^2 units).

It is relevant to recall that the expectation or mean of a continuous random variable Z is

$$E(Z) = \bar{Z} = \int_{-\infty}^{+\infty} Z p(Z)\, dZ \quad (\text{in } Z \text{ units}) \tag{3.53}$$

The variance of Z is defined as the expectation of the squared standard deviation of Z from the expectation of Z, or

$$\sigma^2 \quad E(Z-\bar{Z})^2 = \int_{-\infty}^{+\infty} (Z-\bar{Z})^2 p(Z)\, dZ = \bar{Z}^2 - (\bar{Z})^2 \quad (\text{in } Z^2 \text{ units}) \tag{3.54}$$

As always, the overscribed bar indicates a time average.

The variance is a measure of the "spread" of the distribution, i.e. the deviation of the random variable from its mean—the smaller the variance, the less the deviation.

A plot of the normal density function $p(Z)$ is shown in Fig. 3.23. The random variable may take any values from $-\infty$ to $+\infty$.

By definition, the product $p(Z)\,dZ$ is the probability that the random variable Z occurs in the interval from Z to $Z + dZ$.

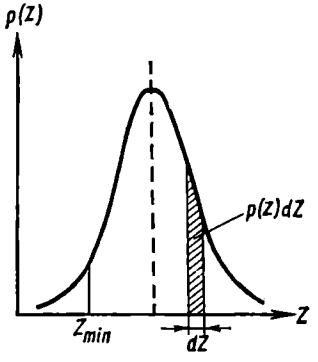


Fig. 3.23. Normal density curve

The probability that Z exceeds some minimum value Z_{min} is given by

$$P(Z_{min}) = \frac{1}{\sigma\sqrt{2\pi}} \int_{Z_{min}}^{\infty} e^{-(Z-a)^2/2\sigma^2}\, dZ \tag{3.55}$$

The relationship $P(Z_{min})$ is the distribution function shown in Fig. 3.24. In the plot, the values of E, in decibels relative to the median value, are plotted as ordinate, and the respective probabilities, or the time during which the specified level is exceeded (as the percentage of the total time) are laid off as abscissa.

For practical calculations, it is convenient to reduce (3.55) to the canonical form of the probability integral

$$P(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-y^2/2}\, dy \tag{3.56}$$

the numerical values of which are given in all texts on probability theory and mathematical reference books.*

Some values of the probability integral are given in Table 3.4.

Table 3.4. SOME VALUES OF THE PROBABILITY INTEGRAL

x	$-\infty$	-1	0	$+1$	$+\infty$
$P(x)$	1	0.84	0.50	0.16	0

* Mathematical reference books often give tables of the function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-y^2/2}\, dy$$

related to $P(x)$ as

$$P(x) = 0.5 - \Phi(x),$$

such that if $x < 0$, it takes the form

$$P(x) = 0.5 + \Phi(|x|)$$

We need not use the general equations, (3.53) and (3.54), each time in order to determine the expectation and variance of a random variable Z. It will suffice to reduce (3.55) to the canonical form of

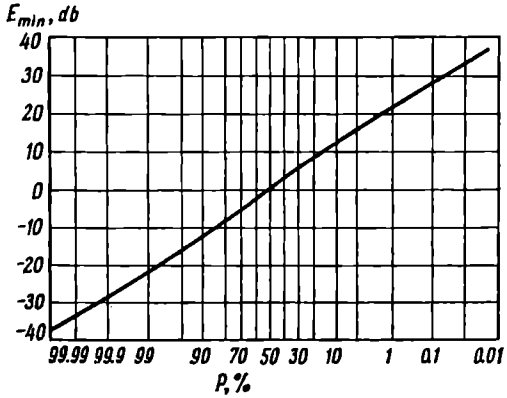


Fig. 3.24. Level distribution function for slow variations in the signal

(3.56) and to find the values of interest from the table of the probability integral. In order to reduce (3.55) to the form of (3.56), we should set

$$y = \frac{Z - a}{\sigma} \tag{3.57}$$

whence

$$dy = dZ/\sigma \tag{3.58}$$

and

$$x = \frac{Z_{min} - a}{\sigma} \tag{3.59}$$

By definition, the median value is that value of Z for which the probability $P(Z_{min}) = 0.5$. Referring to Table 3.4, we find that for $P(x) = 0.5$, $x = 0$. Substituting this value in (3.59) and writing Z_{med} instead of Z_{min}, we get

$$Z_{med} = a \tag{3.60}$$

which explains the physical meaning of a.

Let us now elucidate the physical meaning of the standard deviation σ. To begin with, we find the probability of the level $Z_{med} + \sigma$ being exceeded. Substituting $Z_{med} + \sigma$ in (3.59) for Z_{min} and noting (3.60), we get

$$x = \frac{Z_{med} + \sigma - a}{\sigma} = \frac{a + \sigma - a}{\sigma} = 1$$

Table 3.4 shows that $x = 1$ corresponds to $P(x) = 0.16$.

In a similar way, we find the probability of the level $Z_{med} - \sigma$ being exceeded. Substituting $Z_{med} - \sigma$ in (3.59), we get

$$x = \frac{Z_{med} - \sigma - a}{\sigma} = -1$$

which corresponds to $P(x) = 0.84$ in Table 3.4.

Thus, the standard deviation is a deviation from the median value such that the probability of this level being exceeded is 16 and 84 per cent, respectively. If σ is small, these probabilities occur already with small deviations about the median value.

Some random variables have a logarithmically normal distribution (including those which cannot take on negative values for fundamental reasons). A. N. Kolmogorov has shown that the logarithmically normal distribution applies to particles in comminution. This distribution also applies, as already noted, to the slow random changes in the signal level in the tropospheric propagation of ultrashort waves. In this case, the random variable is the r.m.s. signal field, that is, a substantially positive quantity. As already noted, it is these quantities that have the logarithmically normal distribution. The quantity whose distribution function is sought is the average value of the r.m.s. field strength over five- or ten-minute intervals. Naturally, averaging smoothes the effect of fading. Yet, it is these field strength values averaged over five- and ten-minute intervals that have the logarithmically normal distribution characterizing the slow random changes in the received signal due to changes in the meteorological conditions.

If the field strength E_{rms} at the receiver has a logarithmically normal distribution, the distribution density function is

$$p(E_{rms}) = \frac{1}{E_{rms}\beta\sqrt{2\pi}} e^{-(\log_{10}E_{rms} - \alpha)^2/2\beta^2} \text{ m/V} \qquad (3.61)$$

By definition, the probability distribution density $p(E_{rms})\, dE_{rms}$ is the probability that E_{rms} occurs in the interval from E_{rms} to $E_{rms} + dE_{rms}$. Substituting the expression for $p(E_{rms})$ from (3.61) and placing E_{rms} under the differential sign gives

$$p(E_{rms})\, dE_{rms} = \frac{1}{\beta\sqrt{2\pi}} e^{-(\log_{10}E_{rms} - \alpha)^2/2\beta^2}\, d(\log_{10}E_{rms}) \qquad (3.62)$$

Equation (3.62) may be interpreted as the probability that $\log_{10}E_{rms}$ will occur in the interval between $\log_{10}E_{rms}$ and $\log_{10}E_{rms} + d(\log_{10}E_{rms})$.

On comparing the expressions for $p(Z)\, dZ$ having a normal distribution [Eq. (3.52)] and for $p(E_{rms})\, dE_{rms}$ having a logarithmically normal distribution [Eq. (3.62)], we can readily note that

$\log_{10} E_{rms}$ in (3.62) does what Z does in (3.52). As E_{rms} varies from zero to infinity, $\log_{10} E_{rms}$ varies from $-\infty$ to $+\infty$ (that is, in the same way as Z). If the logarithm of a variable has a normal distribution, the variable itself has a logarithmically normal distribution.

Another difference between (3.52) and (3.62) is that Z having a normal distribution is a dimensional quantity while $\log_{10} E_{rms}$ is dimensionless. Because of this, it is important to specify the units for the field strength in advance.

As far as radio propagation is concerned, it is customary to express field strength in decibels referred to 1 microvolt per metre. Setting

$$\alpha = \log_{10} a, \quad \text{and} \quad \beta = \log_{10} \sigma$$

in (3.61), multiplying and dividing by 20 the right-hand side of (3.62), and also multiplying by 20 the numerator and denominator of the exponent, we may write

$$p(E_{rms,\,db})\, d(E_{rms,\,db}) = \frac{1}{\sigma_{db}\sqrt{2\pi}}\, e^{-(E_{rms,\,db}-a_{db})^2/2\sigma_{db}^2}\, d(E_{rms,\,db}) \tag{3.63}$$

where, as usual,

$$E_{rms,\,db} = 20 \log_{10} E_{rms}$$

$$a_{db} = 20 \log_{10} a$$

and

$$\sigma_{db} = 20 \log_{10} \sigma$$

The probability that the field strength at the receiver will exceed a certain minimum value $E_{min,\,db}$ is given by

$$P(E_{min,\,db}) = \frac{1}{\sigma_{db}\sqrt{2\pi}} \int_{E_{min,\,db}}^{\infty} e^{-(E_{rms,\,db}-a_{db})^2/2\sigma_{db}^2}\, d(E_{rms,\,db}) \tag{3.64}$$

Let us elucidate the physical meaning of a and σ.

From a comparison of (3.64) with the canonical form of the probability integral [Eq. (3.56)], we find that

$$y = \frac{E_{rms,\,db} - a_{db}}{\sigma_{db}} \tag{3.65}$$

and

$$x = \frac{E_{min,\,db} - a_{db}}{\sigma_{db}} \tag{3.66}$$

From Table 3.4 it follows that for the median value $x = 0$. Setting $E_{min,\,db} = E_{med,\,db}$ from (3.66) we get

$$a_{db} = E_{med,\,db} \tag{3.67}$$

In other words, a_{db} is the median value of field strength in decibels referred to 1 microvolt per metre.

Let us now find the probability that the level $E_{med,\,db} + \sigma_{db}$ will be exceeded. Substituting this expression for $E_{min,\,db}$ in (3.66) and writing a_{db} in accordance with (3.67), we have

$$x = \frac{E_{med,\,db} + \sigma_{db} - E_{med,\,db}}{\sigma_{db}} = 1$$

As follows from Table 3.4, the corresponding probability is 16 per cent. In a similar way, we find that the probability that the level $E_{med,\,db} - \sigma_{db}$ will be exceeded is 84 per cent. Thus, σ_{db} is the standard deviation in decibels. In ordinary notation, $E_{min,\,db} = E_{med,\,db} + \sigma_{db}$ is equivalent to

$$E_{min} = E_{med,\,\mu\,V/m} \times 10^{\sigma_{db}/20}\ mV/m$$

Observations of slow random changes in the signal level in tropospheric scatter propagation have shown that the standard deviation is about 12 decibels.

Expressing the probability in per cent and E_{min} in decibels relative to the median value of field strength, the dependence of E_{min} on P may be plotted as shown in Fig. 3.24, where the scale for P (laid off as abscissa in per cent) is chosen such that the logarithmically normal distribution function will be represented by a straight line. The convenience of this representation lies in the fact that one can plot the observed probabilities, draw a curve through the points obtained, and say at first glance how much the distribution function differs from the logarithmically normal distribution.

From reference to the plot of Fig. 3.24, based on observations over a year, it is seen that the received signal varied within ± 30 decibels about the median value during 99.9 per cent of the observation time.

Rapid changes in the signal level, that is, fading proper, are described by the Rayleigh distribution. Theoretical studies into the matter show that this distribution applies to the resultant signal amplitude due to the interference of n sinewaves of arbitrary amplitudes and random phases, where n is an integer greater than four.

In this case, the distribution density is given by

$$p(E_{rms}) = \frac{2E_{rms}}{\overline{(E^2_{rms})}}\, e^{-E^2_{rms}/\overline{(E^2_{rms})}}\ m/V \qquad (3.68)$$

where E_{rms} is the rms value of field strength. This is the rms value of the r.f. signal whose amplitude (and, consequently, the rms value) is changing continually with time due to fading. $\overline{(E^2_{rms})}$ is the mean square value over the observation time T which is usually taken

to be 5 to 15 minutes for tropospheric propagation. In terms of probability theory, $(\overline{E_{rms}^2})$ is the second-order moment of E_{rms}. A plot of the distribution density function is shown in Fig. 3.25.

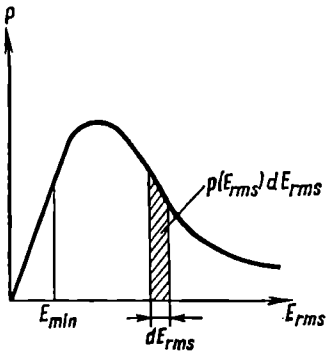


Fig. 3.25. Plot of the Rayleigh distribution density function

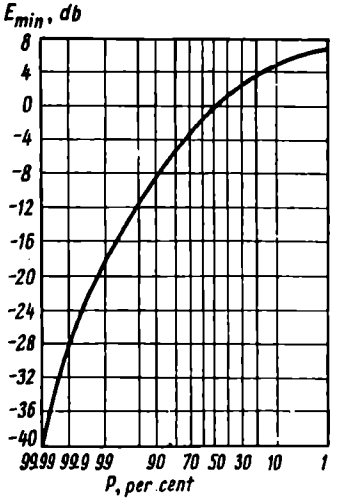


Fig. 3.26. Plot of the distribution density function for fading

E_{rms} is laid off as abscissa, and the respective distribution density, as ordinate.

The probability that the signal level will exceed a certain minimum value E_{min} is given by

$$P(E_{min}) = \frac{2}{(\overline{E_{rms}^2})} \int_{E_{min}}^{\infty} E_{rms} e^{-E_{rms}^2/\overline{E_{rms}^2}} dE_{rms} = e^{-E_{min}^2/\overline{E_{rms}^2}} \qquad (3.69)$$

A plot of the distribution function according to (3.69) is shown in Fig. 3.26. $P(E_{min})$ in per cent is laid off as abscissa, using the same scale as in Fig. 3.24, and E_{min} in decibels referred to the median value, as ordinate.

Using (3.69), we can readily express the median value in terms of the rms value. Setting $P(E_{min}) = 0.5$, we get

$$E_{med}^2 = 0.69\,(\overline{E_{rms}^2})\ \text{V}^2/\text{m}^2 \qquad (3.70)$$

Substituting E_{med}^2 for $(\overline{E_{rms}^2})$ in (3.69) and noting (3.70), the distribution density function may be given a more practical form

$$p(E_{rms}) = \frac{1.38E_{rms}}{E_{med}^2} e^{-0.69E_{rms}^2/E_{med}^2} \qquad (3.71)$$

In pure form, the Rayleigh distribution is a rare occurrence. When the field at the receiver is due to the interference of a wave of a fixed amplitude and phase and a multitude of waves or arbitrary amplitudes and random phases (the latter make up a Rayleigh ensemble), we have what is known as the *non-central Rayleigh* or the *Rician* density function [52].

Assigning an amplitude of unity to the constant component and designating the ratio of the mean square value of the sum of the components making up the Rayleigh ensemble to the rms value of the constant component as k^2, the non-central Rayleigh distribution function may be written as

$$P(r_{min}) = \frac{2}{k^2} \int_{r_{min}}^{\infty} r e^{-(1+r^2)/k^2} I_0(2r/k^2)\, dr \qquad (3.72)$$

where r is the amplitude of the total field (referred to amplitude of unity), and

$$I_0(x) = J_0(ix)$$

where $J_0(ix)$ is the Bessel function of the first kind of order zero of a purely imaginary argument.

The non-central Rayleigh distribution functions for various values of k, that is, the ratios of the power of random components (the Rayleigh ensemble) to that of the constant component in decibels are shown in Fig. 3.27. As always, the probabilities are laid off as abscissa, and the levels in decibels referred to the constant component, exceeded over a certain time, as ordinate. The scale along the X-axis is different from that used in Figs. 3.24 and 3.26. More specifically, the scale used for the probabilities in Fig. 3.27 is such that the Rayleigh distribution will be represented by a straight line. The plot of Fig. 3.27 shows that with large values of k the resultant distribution approaches the Rayleigh distribution, but, in contrast to it, has low (negative) values of k_{db}. With k tending to minus infinity, the resultant amplitude degenerates into a vector of fixed amplitude, which is the constant component of the combining waves.

All forms of fading, irrespective of wavelength and mode of propagation, show time, spatial, and frequency selectivity.

By "time selectivity" of fading is meant the following. Let the time dependence of a field subject to fading be described by a random function $f(t)$ and let the function $f(t + \tau)$ describe this time dependence for a time $(t + \tau)$. The correlation that exists between the events occurring at the time t and $(t + \tau)$ is described by what is

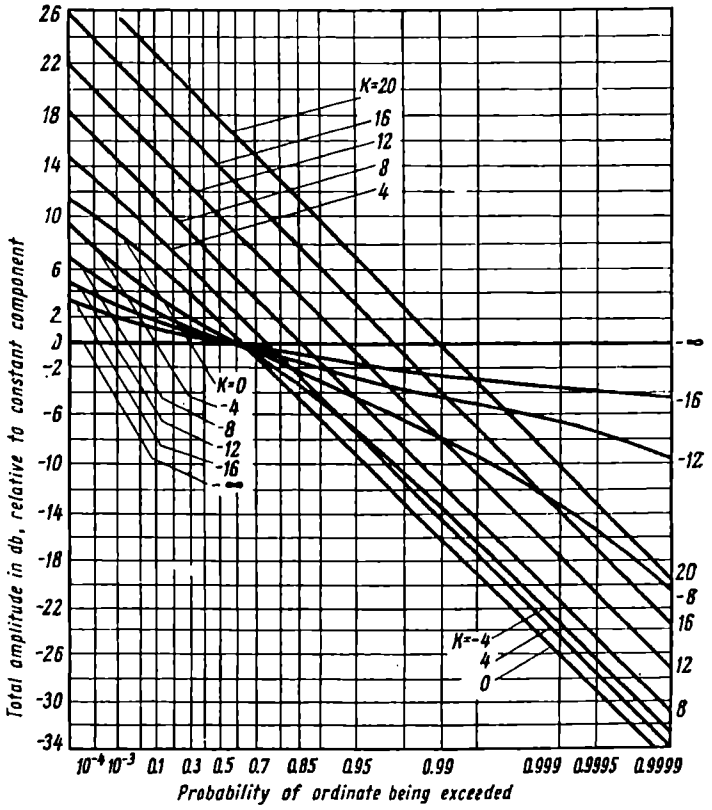


Fig. 3.27. Plot of the non-central Rayleigh or Rician density function

known as the auto-correlation function, $R(\tau)$, defined as

$$R(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} f(t)\, f(t+\tau)\, dt = \overline{f(t)\, f(t+\tau)} \qquad (3.73)$$

where the overscribed bar indicates a time average, as always.

From (3.73) it follows that the auto-correlation function is the product of the event at time t and the event at time $(t + \tau)$, averaged over a period of time long enough to smooth out instantaneous fluctuations. After integration, the variable t naturally disappears, so that the independent variable of the auto-correlation function is the time shift τ.

The auto-correlation function gives information about how the value of $f(t)$ at one time can be used to predict $f(t)$ at a time in

terval τ later. With τ tending to zero, the auto-correlation function tends to a maximum:

$$R(0) = \overline{f_2(t)}$$

As τ increases, $R(\tau)$ decreases monotonically.

In practical calculations, it is more convenient to use the auto-correlation coefficient defined as

$$\rho(\tau) = \frac{\overline{[f(t) - \overline{f(t)}]\,[f(t+\tau) - \overline{f(t)}]}}{\overline{f^2(t)} - [\overline{f(t)}]^2} \tag{3.74}$$

which is unity at $\tau = 0$, and tends to zero as τ increases.

Under actual conditions, the functions $f(t)$ and $f(t+\tau)$ may be regarded as statistically independent already at some finite value of τ. And this is the essence of time selectivity as regards fading.

Spatial selectivity consists in that when the signal is simultaneously received at points spaced a distance l apart, its fading at each point grows increasingly more independent of that at the other points as l increases. Beginning with some critical separation between the aerials, l_{crit}, fading at one aerial is practically independent of that at any other aerial. The explanation lies in the fact that the radio waves arriving at the separated aerials are generated under somewhat different conditions, although they are within the limits of a common scattering volume.

Let $f_1(t)$ be a random function describing fading at one aerial, and $f_2(t)$, that at the other. The relation between them can be expressed in terms of the cross-correlation function which is defined as the average product of one signal $f_1(t)$ and the second signal $f_2(t)$ at time t. As with the auto-correlation function, use is often made of the cross-correlation coefficient:

$$R_{1,2}(l) = \frac{\overline{[f_1(t) - \overline{f_1(t)}]\,[f_2(t) - \overline{f_2(t)}]}}{\overline{f_1^2(t)} - \overline{(f_1(t))^2}} \tag{3.75}$$

At $l = 0$, $R_{1,2}(0) = 1$, while with increase of l, it monotonically tends to zero.

Experiments have shown that in tropospheric propagation with the aerials spaced a distance $l = 100\lambda$ apart, the cross-correlation coefficient is very small, and fading at one of the spaced aerials is independent of that at the other.

For the same reason, in the simultaneous reception of two frequencies emitted by the same transmitter and differing by Δf, the fading at one frequency will be the more independent of that at the other, the greater this frequency separation. Experiments have shown that for a working frequency of 2.3 gigahertz and a frequency separation of 2 megahertz the cross-correlation coefficient is 0.4, while at $\Delta f = 4$ megahertz, it reduces to about 0.1 [53].

The spatial and frequency selectivity of fading has for long been used in radio communication to minimize the signal distortion caused by fading.

Let S_1 be the probability that in reception with one aerial the signal will fall below a threshold value E_{min} (the one at which reliable reception of signals is still possible). The quantity S_1 is related to $P(E_{min})$ as

$$S_1 = 1 - P(E_{min}) \tag{3.76}$$

If, for example, $S_1 = 10$ per cent, then the signal will be reliably received during 90 per cent of the working time of the radio link.

In reception with two spaced aerials the probability that the signal will fall below the threshold value simultaneously at both aerials can be determined by a theorem used in probability theory. The theorem states that if S_1, S_2, etc. are the probabilities of independent events, the probability that these events will occur simultaneously is equal to the product of the probabilities of these events.

In our case it is assumed that the aerials are spaced apart a distance such that fading at each will be independent of that at the others. With n separated aerials, the probability that the reception will be poor is given by

$$S_n = S_1^n \tag{3.77}$$

Method for the combination of signals in reception with spaced (diversity) aerials and with the use of frequency diversity (known as diversity reception) may be found in texts on radio reception and radio receivers.

Fig. 3.28 shows curves representing the distribution of probabilities that the signal will exceed a certain minimum value (in decibels referred to the median signal value in reception with one aerial) for two, three and up to eight spaced aerials. As is seen, fading can be effectively controlled by use of space diversity, frequency diversity, or their combination. This is especially true of reception with two aerials in comparison with that using one aerial. Use of a greater number of spaced aerials is advantageous too, but to a more limited extent than when two aerials are used instead of one. It is relevant to note that space diversity cannot be used to control slow changes in the received signal.

Another distinction of fading is its rate. As already noted, the rate of fading may in a first approximation be taken to be equal to the average duration of fading. A more accurate picture is given by the distribution of the average duration. Still more convenient for many cases is the frequency spectrum of fading, that is, the frequency spectrum of the signal envelope.

We have seen that the spatial spectrum $F(K)$ of tropospheric fluctuations is related to the spatial correlation function $C(\rho)$ by

a Fourier transform, (3.24). In much the same way, the frequency spectrum $F(\omega)$ of fading (in the time domain) is related to the auto-

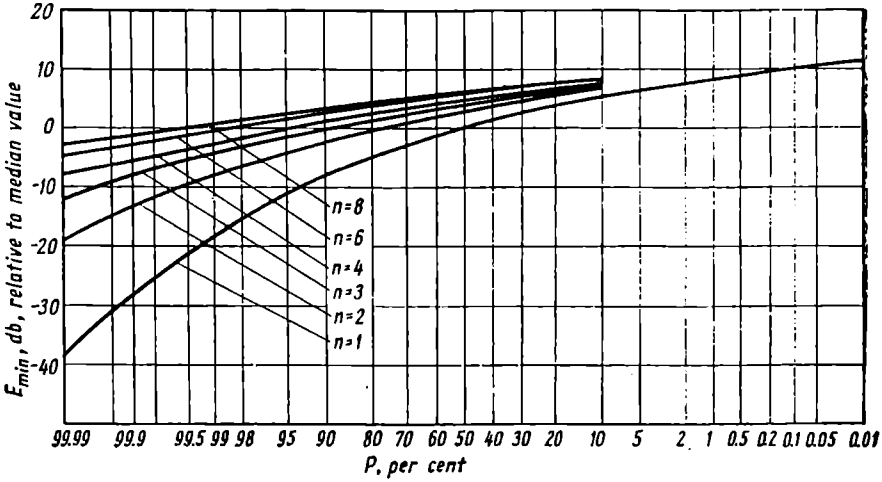


Fig. 3.28. Curves representing the distribution of probabilities that the signal will exceed a certain minimum value (in decibels referred to the median signal value in reception with one aerial) for two, three and up to eight spaced aerials

correlation coefficient $\rho(\tau)$ (3.74). Applying the Fourier transform to the auto-correlation function gives

$$F(\omega) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \rho(\tau)\, e^{-i\omega\tau}\, d\tau \text{ (seconds)} \qquad (3.78)$$

3.12. Seasonal Variations in the Signal Level in Tropospheric Propagation of Ultra-short Waves. The Effect of Climatic Conditions

Outwardly, the seasons of the year affect tropospheric scatter propagation in that the signal level along propagation paths in the Northern Hemisphere is higher in summer than it is in winter. At temperate latitudes, the seasonal variations in the signal level may be 10 to 12 decibels.

Seasonal variations in field strength are shown well in Fig. 3.29. The gradual curve in the upper plot represents variations in the average field strength over a three-year period, as measured at 104.5 megahertz along a propagation path of 282 kilometres [54]. The lower plot represents variations in the average refractive index N_e

near the surface a half-way along the same path over the same period. As is seen, the two curves have about the same shape. The close relationship between the average refractive index N_e near the surface and the scattered signal level may be explained as follows. As winter

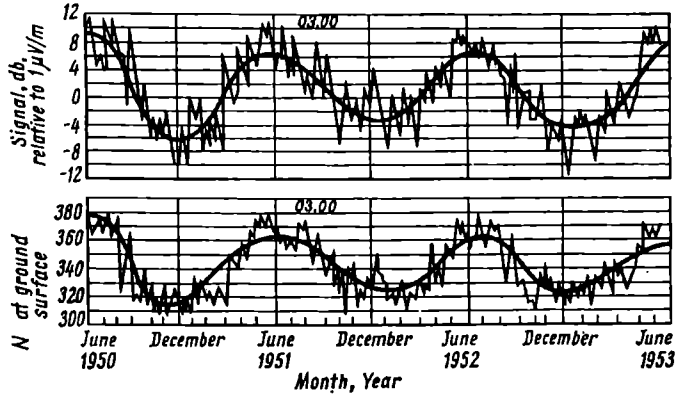


Fig. 3.29. Seasonal variations in field strength

gives way to summer, the average temperature and total humidity of the air increase, and N_e also increases. At the same time, the refractive index near the upper boundary of the troposphere remains

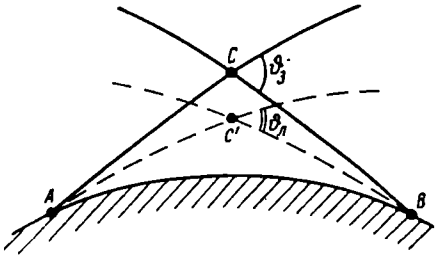


Fig. 3.30. Ray traces for tropospheric scatter propagation in winter (full lines) and summer (broken lines)

almost unchanged, and the gradient dN/dh decreases (the gradient is always negative; it is its absolute value that increases).

From Eq. (3.10a) it follows that a decrease in the gradient entails a decrease in the radius of the ray's curvature in the troposphere. Fig. 3.30 shows ray traces for tropospheric scatter propagation in winter (the full lines) and summer (the broken lines). While in winter the lower point of the common scattering volume is at C, in summer the greater bending of rays shifts it to C', and the sca-

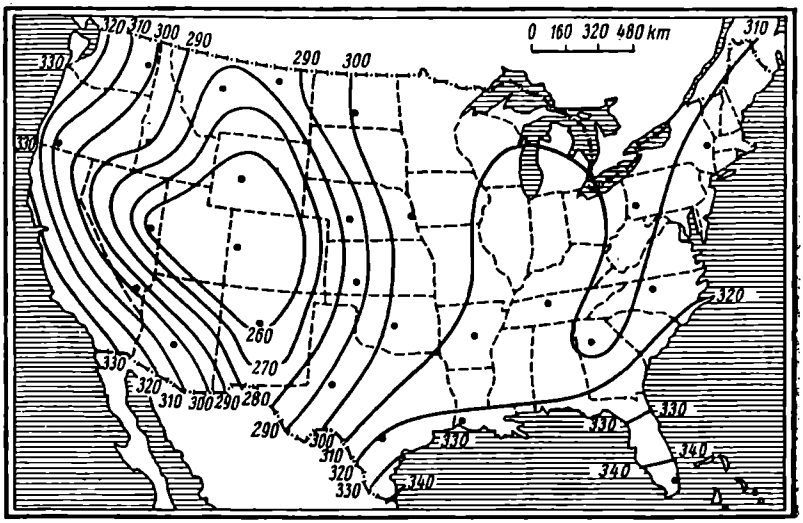


Fig. 3.31. Isopleths of refractive index for the United States (winter)

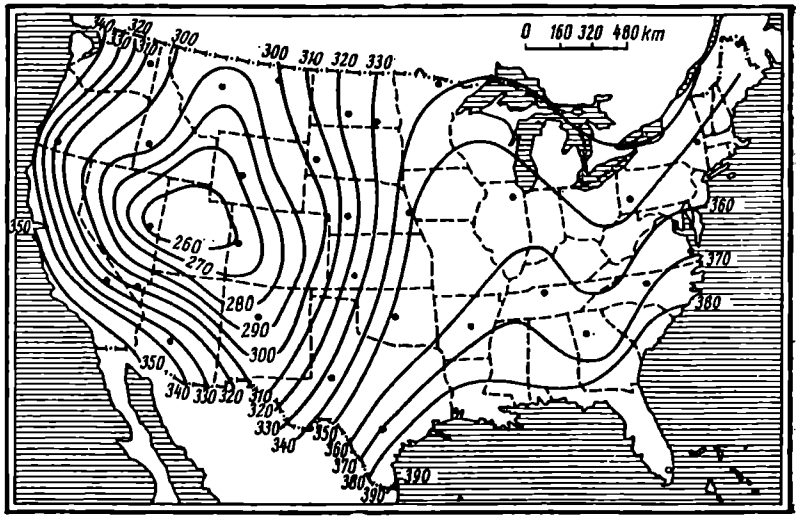


Fig. 3.32. Isopleths of refractive index for the United States (summer)

ttering angle θ is decreased. This entails further effects due to two factors. Since, according to the most widely used theories of scattering, the effective scattering cross-section σ is inversely proportional to the fifth power of θ, even an insignificant decrease in the scattering angle is bound to produce an increase in the scattered field. On the other hand, the mean square variation in refractive index (and the vertical gradient of the refractive index) always increases on moving closer to the surface and this causes a further increase in the signal level. Aiding each other, the two factors produce a marked increase in the signal level in summer as compared with winter.

A warm, maritime climate favours tropospheric scattering, while a cold, dry climate weakens the scattered field. A general idea about the intensity of tropospheric scattering in a specific area of the globe can be gleaned from a map of that area, bearing isopleths of refractive index N_e near the surface, obtained for winter and summer. Examples of such maps, compiled for the United States, are shown in Fig. 3.31 (winter) and Fig. 3.32 (summer). As is seen, in summer the most favourable conditions for tropospheric scatter propagation occur near the Gulf of Mexico, while in winter the worst conditions for tropospheric scatter propagation are to be expected in the mountains (the State of Colorado).

3.13. Approximate Calculations of Tropospheric Radio Links

Using field-strength measurements in the ultra-short wave band at large distances from the transmitter under conditions excluding

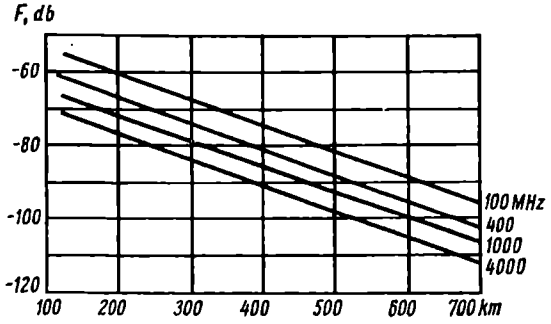


Fig. 3.33. Median values of the attenuation function plotted against distance for winter

tropospheric ducting, it has been possible to establish the relationship between attenuation function and distance for the various frequencies in the range from 100 megahertz to 4 gigahertz. A plot of this relationship appears in Fig. 3.33. It uses the median values

(see § 3.11) of the attenuation function for winter at the temperate latitudes of the Northern Hemisphere, with radio waves propagated over land. This qualification is essential because, in view of the preceding section, the scattered field at the receiver is strongly dependent on the average value of the refractive index near the surface N_e.

In the Northern Hemisphere, as winter gives way to summer, the temperature and humidity of the air increase, and the refractive index near the surface also increases. While at 40° Lat. N. the average refractive index in winter is 300 N-units, it rises to 380 N-units in summer (see Fig. 3.29). As is shown in § 3.12, this increase in N_e leads to a stronger field. For the same reason, engineering calculations should be based on the attenuation function-vs.-distance plot for winter. If communication is reliable in winter, it will be the more so in summer.

Using the family of curves shown in Fig. 3.33 one can readily find the median values of the scattered field at the receiver at a specified distance from the transmitter. Fig. 3.33 shows that the attenuation function is to a certain degree dependent on the emitted frequency. Experimental data check well with Villars-Weiskopf's theory based, in a sense, on Kolmogorov's theory of the local isotropy of turbulence.

As is shown in § 3.9, the power at the receiver input is given by (3.33) which is written again for convenience

$$P_2 = \frac{P_1 G_1 G_2 \lambda^2 F^2}{(4\pi r)^2} \text{ (W)} \tag{3.79}$$

where F is the field-strength attenuation function related to F_{db} (which can be found from the plot of Fig. 3.33) as

$$F = 10^{F_{db}/20}$$

Let the noise at the receiver* and the receiving equipment used be such that a specified form of transmission (multi-channel telephony, television, etc.) may be received reliably, that is, with the power at the receiver exceeding some threshold value of P_2. Substituting this value in (3.79) will give the values of P_1, G_1, and G_2 (or P_1, if G_1 and G_2 are specified in advance) which will secure the desired input power.

However, the parameters of the radio link so chosen cannot provide for reliable operation yet, because the plot of Fig. 3.33 gives only the median value of the attenuation function, and then only for winter. In other words, the power at the receiver input will

* Practical methods for finding the noise level at the receiver are discussed in Chapter 6.

exceed the specified value only during 50 per cent of the operating time of the link in winter. That is, the reliability of communication under the circumstances will only be 50 per cent. To secure more reliable communication, it is important to increase the power of the transmitter (or the gains of the transmitting and receiving aerials). How great this increase must be will depend on the desired reliability

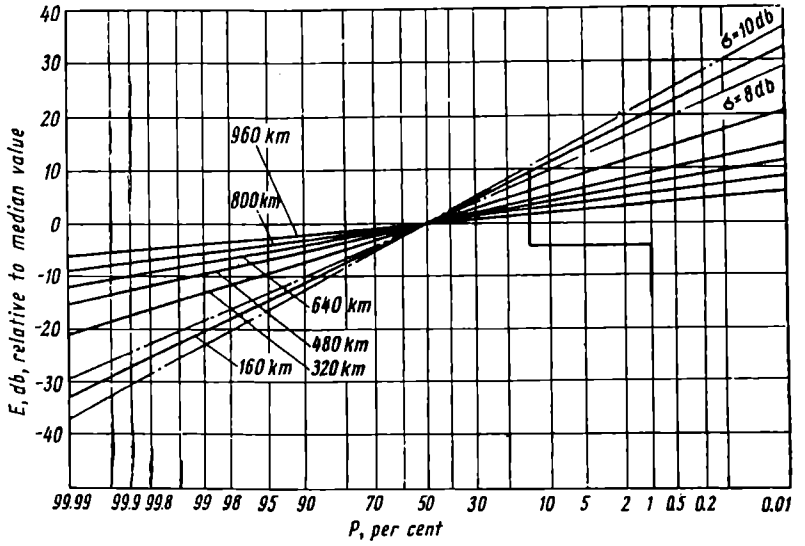


Fig. 3.34. Distribution of probabilities that some minimum signal will be exceeded (in decibels relative to the median value), describing slow variations in the received signal

of communication and the number of diversity aerials used. It should also be taken into account that the field at the receiver is subject to both slow changes (due to changing weather conditions) and fading (owing to the fact that a multitude of interfering rays reach the receiver). Therefore, the final choice of P_1, G_1 and G_2 should provide for the compensation of both slow and fast changes in the field strength.

The margin of transmitted power necessary to compensate for slow changes in field strength can be determined from Fig. 3.34 where the desired reliability of communication (in per cent) is laid off as abscissa, and the ratio (in decibels) of the median field strength to the hourly average exceeded during the per cent time laid off along the X-axis, as ordinate*. The curves of Fig. 3.34 were

* Some authors believe that better results are obtained with an average over a few minutes.

obtained experimentally [55]. It is seen that the depth of slow changes in the field strength decreases with increasing propagation path length. From the plot of Fig. 3.34 it is an easy matter to determine how much the transmitter power (or, more accurately, the product $P_1G_1G_2$) must be increased so that the hourly average field strength will exceed the desired value 99.9 per cent of the time.

However, there is one more limitation to be overcome. The point is that fading may cause the instantaneous field strength to deviate from the hourly average considerably. In order to find a further increase in transmitter power (to take care of fading), reference should be made to Fig. 3.28.

From Fig. 3.28, it follows that in reception with one aerial the field exceeded 99.9 per cent of the operating time is 18 decibels below the median value. Consequently, in order that the threshold signal at the receiver be exceeded 99.9 per cent of the time, the transmitter power P_1 (or the product $P_1G_1G_2$) must be increased by another 18 decibels. If both space and frequency diversity are used at the same time, the requisite increase in the product $P_1G_1G_2$ may similarly be found from the curves of Fig. 3.28, corresponding to the number of diversity aerials used.

Example 3.2. Determine the transmitter power necessary to secure a communication reliability of 99.9 per cent over a propagation path 400 km long, for a frequency of 800 MHz ($\lambda = 37.5$ cm), and aerial gains of $G_{1,\,db} = G_{2,\,db} = 40$ db ($G_1 = G_2 = 10^4$). Fading is controlled by dual frequency diversity and dual space diversity.

Solution: Suppose that with the noise level existing at the point of reception, a given type of receiving equipment, and a given form of transmission the threshold power at the receiver input should be

$$P_{2,\,th} = 5 \times 10^{-13} \text{ W}$$

Referring to the plot of Fig. 3.33, the median value of the attenuation function for winter is $F = -81$ decibels or $F = \frac{1}{1.1 \times 10^4}$. From the plot of Fig. 3.34, we find that for a propagation path 400 kilometres long the margin of transmitted power to compensate for slow changes in the field and to secure a reliability of 99.9 per cent must be 15 decibels.

From the plot of Fig. 3.28 we find that reception with four diversity aerials calls for a transmitted-power margin of 5.5 decibels so as to compensate for fading and secure a reliability of 99.9 per cent.

Substituting the numerical values thus obtained in (3.79), the power necessary to secure a 50-per cent reliable reception with one aerial in winter will be

$$P_1 = \frac{(4\pi r)^2 P_{2,\,th}}{G_1^2 F^2} = 15 \text{ W}$$

For quadruple diversity reception with an overall reliability of 99.9 per cent, the transmitter power must be raised by $15 + 5.5 = 20.5$ decibels, that is, 110 times. Thus,

$$P_1 = 15 \times 110 = 1650 \text{ W}$$

In conclusion, it should be stressed again that tropospheric scatter affects only ultra-short waves. As to trans-horizon propagation of short, medium, and long waves, other factors come into play, namely ionospheric diffraction and reflection. On the other hand, these factors do not affect ultra-short waves.

The discovery of tropospheric scatter propagation for ultra-short waves has necessitated a revision of the earlier views under which they were regarded as suitable only for short-distance communications. Experiments and pilot tropospheric-scatter links have shown that with high-power transmitters (20 to 50 kilowatts) and pencil-beam transmitting and receiving aerials (with a diameter of 20 metres and more), it is possible to provide reliable communication at wavelengths from one metre to a few centimetres over distances up to a thousand kilometres, with an undistorted bandwidth of about five megahertz. In other words, tropospheric scatter links may well be used for multi-channel telephony and television. Broad-band tropospheric-scatter links may cover distances up to 300 or 400 kilometres.

3.14. Attenuation of Radio Waves in the Troposphere

So far, the effects of the troposphere on radio propagation have been dealt with from the view-point of its inhomogeneity resulting in the bending of radio rays (atmospheric refraction) in some cases, and in the scattering of ultra-short waves in others (due to local irregularities). In all cases, however, it has been tacitly assumed that the troposphere is fairly transparent to radio waves, that is, radio waves experience no attenuation as they are propagated through it.

Practical operation of radio links within a rather wide range of wavelengths over nearly seventy years has shown that waves longer than ten centimetres do not experience an appreciable attenuation in the troposphere in the face of torrential rain, fog, snow, hail, clouds or other meteorological factors at work in the lower troposphere. In contrast, waves shorter than ten centimetres are subject to attenuation which, under certain conditions, may be so marked that radio communication is disrupted.

The problem of radio attenuation in the troposphere has of late grown in importance in connection with advances in the use of mil-

limetric waves and the prospects of using submillimetric wavelengths for radio communication. Accordingly, it is important to study radio attenuation in the troposphere over a fairly wide frequency range from 3 gigahertz ($\lambda = 10$ centimetres) to 1,000 terahertz ($\lambda = 0.3$ micron) which encompasses centimetric and millimetric waves, infra-red, visible and ultra-violet rays.

Attenuation of radio waves in the troposphere may be caused by four factors, namely (1) absorption by precipitation particles — rain, fog, hail, and snow; (2) absorption by molecules; (3) scattering by molecules and collections of molecules, notably by a light haze; and (4) absorption by solids (dust, smoke, etc.) appearing in the troposphere as a dry haze.

Attenuation by Precipitation Particles. Two different physical factors causing attenuation by precipitation particles may be named. To begin with, each water droplet may be treated as an imperfect conductor in which the advancing radio wave induces displacement currents. The density of these currents is considerable, because the dielectric constant of water is about 80 times that of air. On the other hand, the density of displacement currents is proportional to frequency, so that heavy currents appear at the highest frequencies in the centimetric and millimetric bands. The resultant absorption of energy by water droplets manifests itself as the attenuation of radio waves.

Besides, the currents induced in the droplets of rain or fog are sources of scattered or secondary radiation. Practically, this scattering results in the attenuation of radio waves in the direction of propagation—instead of being propagated in the right direction, the waves are partly caused to scatter in all directions.

Water droplets in fog range from two to sixty microns in radius. The larger droplets are so heavy that they are no longer capable of floating in the atmosphere and fall out as rain. With positive temperatures of the air, most droplets measure from five to fifteen microns, while at negative temperatures the respective figures are from two to five microns. The number of droplets per cubic metre of air ranges from five to a hundred for light fog and from 500 to 600 for dense fog. A very important characteristic of fog is its liquid water content, that is, the amount of condensed moisture in grams per cubic metre. In a light fog, the visibility is one kilometre; in a dense fog the visibility shrinks to a few metres. Clouds are the same fog. The basic characteristics of fog varying in density are listed in Table 3.5.

Rain drops have a radius of over sixty microns, the upper limit being seven millimetres. Commonly, the radius of rain drops is from 0.25 to two millimetres. The drop sizes and liquid-water contents for rain varying in severity are listed in Table 3.6.

Table 3.5. BASIC CHARACTERISTICS OF FOG

Severity of fog	Average droplet size, microns	Droplets per cm³	Liquid content, g/m³
Light	5	60	0.03
Medium	5	600	0.3
Dense	10	600	2.3

In general, the liquid-water content of rain is only slightly greater than that of fog. For a torrential rain, as an example, the liquid-water content is only 2.5 times that for a dense fog, while that of a drizzling rain is only three times that of a light fog. On the other hand, the liquid-water content of rain and fog is considerably lower

Table 3.6. BASIC CHARACTERISTICS OF RAIN

Rain	Rate of precipitation, mm/hr	Drop radius, mm	Drops per m³	Average distance between drops, cm	Liquid-water content, g/m³
Drizzling	0.25	0.1	—	—	0.092
Slight	1	0.225	—	—	0.14
Moderate	4	0.5	530	12	0.28
Heavy	15	0.75	450	—	0.83
Very heavy	40	1	—	—	1.9
Torrential	100	1.5-2.5	400	14	5.4

than that of an atmosphere laden with vapours. It has been measured that at 25°C the liquid-water content of a moisture-laden atmosphere is 23 grams per cubic metre, which is five times that of a torrential rain.

The attenuation can be described either by the factor

$$e^{-\delta l(\text{km})}$$

where δ is the loss per unit length (one kilometre) of the propagation path, and l is the total distance covered, or by the factor

$$10^{-\delta(\text{db/km})l(\text{km})/20}$$

the attenuations in both cases being in terms of decibel loss per kilometre.

The two factors describe the attenuation of the field due to an advancing radio wave. If the distance covered by the wave is r,

ard the rain area extends over l km of this total path, the field is given by

$$E_{rms} = \frac{173\sqrt{P_{1\,\mathrm{kW}}G_1}}{r_{\mathrm{km}}} e^{-\delta l_{\mathrm{km}}} \text{ mV/m} \qquad (3.80)$$

In the centimetric and millimetric wave bands, the attenuation experienced by radio waves in rain and fog is the result of energy absorption (dissipated as heat) and scattering.

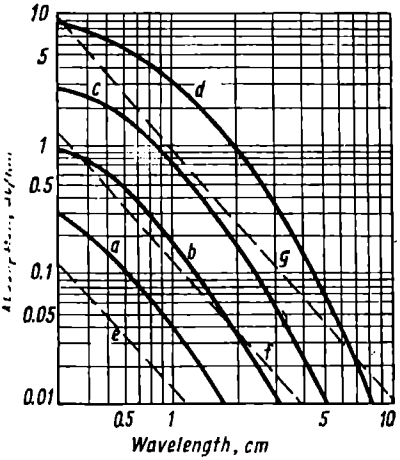


Fig. 3.35. Absorption coefficient versus frequency for rain and fog

In a heavy rain, waves with a length of 3 centimetres will experience an attenuation of 8×10^{-3} per kilometre, as against 1.1×10^{-3} per kilometre without rain. The attenuation increases with decreasing wavelength.

Additional attenuation may be due to the fact that in some cases radio waves may run into and be reflected from well-defined clouds or rain streaks.

Experimental data are most exhaustive as regards the attenuation of waves at $\lambda = 1.25$ centimetres. The observations have been carried out to cover the most commonly encounteed rates of rainfall, namely from 1 to 88 millimetres per hour. These experiments have established a linear dependence of the decibel los, δ, on rainfall rate. The absolute value of δ has been found to be 2.64×10^{-2} per kilometre per millimetre of rainfall per hour, which corresponds to a loss of 0.25 decibel per kilometre per millimtre per hour.

Fig. 3.35 shows a plot relating the attenuation in decibels per kiometre to wavelength in the centimetric and millimetric bands urder conditions of rain and fog [34]. The full lines represent the atenuation by rain, and the broken lines, by fog. The letters at th curves stand as follows (see Tables 3.5 and 3.6): *a*, drizzling ran, 0.25 mm/hr; *b*, slight rain, 1 mm/hr; *c*, moderate rain, 4 mm/hr; *d*, heavy rain, 15 mm/hr; *e*, light fog, 0.03 grams of liquid wter per cubic metre (visibility, about 600 metres); *f*, medium fog, 0. grams of liquid water per cubic metre (visibility, about 120 metrs); *g*, dense fog, 2.3 grams of liquid water per cubic metre (visibility, about 30 metres).

Example 3.3. Determine the field due to a wave freely propagated in a fog (with a visibility of about 30 m) at a distance of 5 km from the transmitter with $P_1 = 1$ W, $G_1 = 600$, and $\lambda = 2$ cm.

Solution: Referring to curve g in Fig. 3.35, for $\lambda = 2$ cm, we find that the attenuation is 0.25 db/km, or

$$F = 10^{\frac{-\delta_{(\text{db/km})}l_{\text{km}}}{20}} = 0.87$$

Substituting it in Eq. (3.80) and noting that in our case $r = l = 5$ km, we get

$$E_{rms} = \frac{173\sqrt{10^{-3} \times 600}}{5} \times 0.87 = 23.4 \text{ mV/m}$$

As follows from the plot of Fig. 3.35, a wavelength of 2 millimetres is attenuated by a heavy rain and a dense fog so much (up

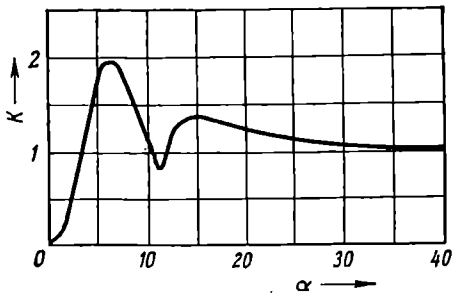


Fig. 3.36. Plot of the Straiton-Hawton formula

to 10 decibels per kilometre) that radio communication is practically impossible already at a distance of 10 kilometres.

In the infra-red and visible regions of the spectrum, energy absorption by water droplets may be ignored, and so there remains only one factor causing the attenuation of waves—scattering. The loss caused by scattering can be evaluated solely from theoretical considerations. Straiton and Hawton [56] have shown that for infra-red rays the attenuation (in decibels per kilometre) is given by

$$\delta_{\text{db/km}} = 4.34 \times 10^5 N_{1/\text{cu. cm}} K 2\pi\rho^2_{\text{cm}} \qquad (3.81)$$

where N is the number of water droplets per cubic centimetre, ρ is the radius, and K is the scattering function found from the plot of Fig. 3.36 where $\alpha = 2\pi\rho/\lambda$ is laid off as abscissa, and λ is the wavelength.

The application of Eq. (3.81) to three rainfall rates, namely 4 millimetres per hour (a moderate rain), 16 millimetres per hour (a heavy

rain), and 100 millimetres per hour (a torrential rain), is illustrated in Fig. 3.37. The wavelength, in the interval from 0.1 micron to one centimetre, is laid off as abscissa. In the interval between 20 and 600 microns the broken curves show that the atmosphere is opaque because of molecular absorption (see below). The dot-dash curves on the right of Fig. 3.37 are taken from Fig. 3.35 and represent the relationships for rainfall rates of 4 and 16 millimetres per hour, determined by experiment.

The Straiton-Hawton equation provides a link between the Rayleigh scattering applicable to cases where the wavelength is many

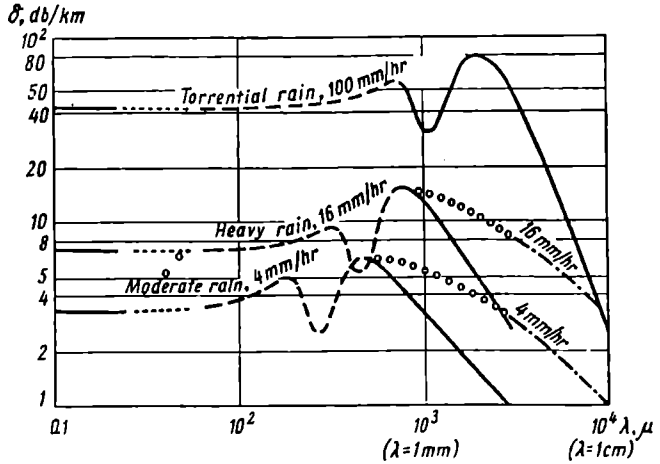


Fig. 3.37. Absorption coefficient as a function of wavelength for three rates of rainfall

times the size of scattering particles (in such cases, the absorption coefficient is inversely proportional to the wavelength raised to fourth power), and Mie's scattering applicable to cases where the size of scattering particles is many times the wavelength. In the latter case, the absorption coefficient is independent of wavelength. In the plot of Fig. 3.37, Mie's scattering holds in the interval between 0.1 and 100 microns, while at wavelengths upwards of 1000 microns one might expect Rayleigh scattering. However, the gradient of the experimental (dash-dot) curves is much smaller. The expected form of curves is shown by the dots in Fig. 3.37.

The application of (3.81) to three fog intensities (light, medium, and dense) is illustrated in Fig. 3.38. As with Fig. 3.37, the curves are discontinuous in the interval from 20 to 600 microns where the classical Rayleigh scattering is observed, and the absorption coefficient is inversely proportional to the wavelength raised to fourth

power. Actually, this is not the case apparently, because there is a marked discrepancy between the theoretical curves and the experimental (dash-dot) curves on the right of the plot, taken from Fig. 3.35. The expected actual form of curves in the interval from 10 to 3 microns is shown by the dots.

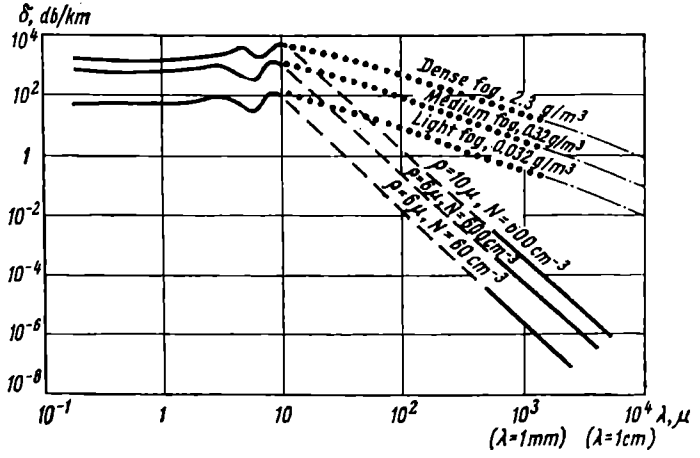


Fig. 3.38. Absorption coefficient as a function of wavelength for three intensities of fog

The conclusion that may be drawn is that the attenuation caused by a heavy rain or a dense fog at wavelengths shorter than 3 centimetres is so strong that radio waves cannot be propagated in the troposphere. There is no ground for expecting that a rain or a fog might have "windows" transparent to radio waves in this band.

Molecular Absorption of Radio Waves. The attenuation of radio waves shorter than 1.5 centimetres may also be due to the interaction of the field of the wave and the molecules of the atmospheric gases. Known as molecular absorption, this phenomenon occurs in the absence of rain, fog or any other precipitation particles. The energy of the advancing wave is wasted to heat the gases, to ionize or excite atoms and molecules, to initiate photo-chemical reactions, and the like. As they absorb radio energy, the atoms and molecules pass from a lower-energy to a higher-energy state. In this process, the major role as regards the atoms is played by the energy of the outer electrons, while for the molecules the decisive factors are the energy of electron oscillation and of molecular rotation. Since the energy states are discrete, these transitions are more or less resonant in character, which results in the *resonant* or *selective absorption* of radio waves.

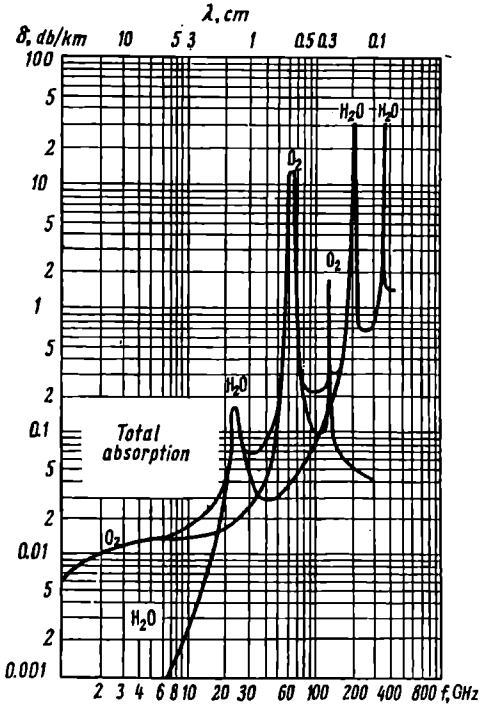


Fig. 3.39. Absorption coefficient for oxygen of the air and water vapour as a function of frequency in the range from 1 GHz (λ = = 30 cm) to 500 GHz (λ = 0.6 mm)

The major atmospheric gases that need to be considered as absorbers are water vapour and oxygen. In the centimetric and millimetric wave bands, the resonant absorption in water vapour occurs at wavelengths of 1.35 centimetres, 1.5 millimetres, and 0.75 millimetre, and in oxygen at wavelengths of 0.5 and 0.25 centimetre.

The plot of Fig. 3.39 (after [57]) relates the absorption in decibels per kilometre to frequency in the range from one gigahertz (λ = = 30 centimetres) to 500 gigahertz (λ = 0.6 millimetre). As is seen, there exist four resonant absorptions in this frequency range. While absorption by oxygen is more or less constant, it is not so with water vapours whose content varies from day to day, according to the humidity of the air. Therefore, it is important to know the humidity of the air to which a given absorption coefficient applies. The plot of Fig. 3.39 has been drawn up for an absolute humidity (or liquid-water content) of 7.75 grams per cubic metre. The absolute

humidity at which saturation occurs (that is, the limiting absolute humidity) depends on the temperature of the air. Its values for different temperatures are quoted in Table 3.7.

Table 3.7. WATER VAPOUR CONTENT, Q, RESULTING IN SATURATION AT VARIOUS TEMPERATURES

t, °C	Q, g/m³	t, °C	Q, g/m³
—10	2.14	10	9.4
—5	3.24	15	12.8
0	4.84	20	17.3
5	6.8	25	23.0

From Table 3.7 it follows that at an elevated temperature, the liquid-water content of the air is higher than that of a torrential rain.

Fig. 3.39 suggests the existence of "windows" in the troposphere relatively transparent to radio waves. One such window, termed the K_α band, lies between the resonant peaks at wavelengths of 0.5 and 1.35 centimetres where the absorption at $\lambda = 0.86$ centimetre drops to 0.5 decibel per kilometre. Several other "windows" might be mentioned.

Waves at optical frequencies (infra-red rays, visible light, and ultra-violet rays) are likewise subject to strong attenuation due to molecular absorption. This is especially true of water vapour whose resonant absorption bands lie so closely together that they form continuous absorption regions.

Yet, in this frequency range, too, there are "windows" transparent to radio waves. To begin with, there is a "window" between 0.4 and 0.85 micron, which covers the entire region of visible light extending from 0.4 to 0.75 micron. This is why we can see the multitude of colours in daylight on clear days and the star-studded sky on clear nights. This and other "windows" in the range from 0.4 to 16 microns are listed in Table 3.8.

To wavelengths longer than 14 microns and up to 1.5 millimetres, the troposphere is opaque, mainly because of the numerous and very strong absorption bands of water vapour.

The plot of Fig. 3.40 relates the transmission coefficient for the radiation that passes through the atmosphere (at normal incidence) to the wavelength in the range 0.9 to 14 microns. The transmission coefficient in the terminology adopted in this book is identical with the power attenuation function, F^2. Since water vapour and

Table 3.8. TROPOSPHERE "WINDOWS" FOR VISIBLE LIGHT AND SHORT INFRA-RED RAYS

Wavelength limits, microns	Transmission coefficient (per cent) for a normally incident ray passing through the entire atmosphere	Wavelength limits, microns	Transmission coefficient (per cent) for a normally incident ray passing through the entire atmosphere
0.4-0.85	100	2.0-2.5	80
0.95-1.05	70	3.2-4.2	90
1.2-1.3	70	4.5-5.2	70
1.5-1.8	80	8.0-13.5	40-80

other atmospheric gases have the highest pressure either near the earth's surface or at a very small height above it (for example, O_3), the plot of Fig. 3.40 gives a true picture of absorption in the

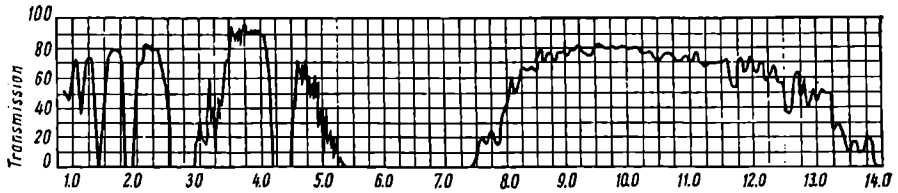


Fig. 3.40. Transmission coefficient of the atmosphere as a function of wavelength in the range from 0.9 to 14 microns

troposphere. The absorption bands with peaks at wavelengths of 0.93, 1.13, 1.40, 1.87, 2.74, 6.3, and 17 microns are due to water vapour. The bands with peaks at wavelengths of 2.7, 4.26 and 15 microns are due to CO_2. The band at $\lambda = 9.5$ microns belongs to ozone.

The general conclusion that can be drawn is that it is possible to maintain radio communication on optical (submillimetric) frequencies in the intervals corresponding to "windows". Naturally, it is presumed that there is neither rain nor fog to interfere with radio propagation.

In space communication, with a radio set carried by an aircraft or spacecraft well above the troposphere, molecular absorption may be neglected to a greater measure or even completely. Nor will there be any attenuation due to precipitation particles for the simple reason that they are non-existent in outer space.

Attenuation of Radio Waves by Molecular Scattering. In dealing with the shorter wavelengths in the optical range, adjacent to visible light, one has to reckon with the scattering caused by the in-

dividual molecules of the atmospheric gases. The linear dimensions of these molecules are of the order of 3×10^{-4} microns, which is much smaller than the wavelengths of visible and, the more so, of infra-red light. Because of this, the loss due to molecular scattering can be evaluated in terms of the classical Rayleigh scattering.

In the notation adopted, the attenuation due to Rayleigh scattering is [56]

$$\delta = \frac{4.34 \times 32\pi^3 \times 10^3}{3N_{1/\text{cu. m}}\lambda_m^4}(n_0 - 1)^2 \text{ db/km} \qquad (3.82)$$

where $N = 2.69 \times 10^{25}$ per cubic metre is Avogadro's number under normal atmospheric conditions, and n_0 is the refractive index at the earth's surface, normally equal to 1.000325.

From Eq. (3.82) it follows that the attenuation is inversely proportional to the fourth power of wavelength, that is, the attenuation due to molecular scattering rapidly decreases with increasing wavelength.

The presence of liquid water markedly affects attenuation, despite the fact that the molecules of water are about the same size as those of O_2 and N_2. The point is that the water vapour present in the troposphere, even in quantities far from saturation, tends to condensate into tiny droplets many times larger in size than the molecules. The nuclei of condensation in such cases are the molecules containing salts of chlorine, sodium, magnesium, and sulphur, which are always present in the troposphere and are strongly hygroscopic. These salts are believed to find their way into the troposphere due to the evaporation of sea-water drops from rough seas. These particles can be carried by the wind over large distances of thousands of kilometres.

At low air humidity, the hygroscopic nuclei are crystals with a mass of 10^{-17} to 10^{-15} gram. As the humidity increases, the crystals turn into droplets of solution, with a radius of 10^{-3} to 10^{-1} micron. At a humidity close to 100 per cent and a considerable concentration of hygroscopic nuclei, a haze appears in the atmosphere, and the attenuation of visible radiation drastically increases while the visibility drops to 10 or even one kilometre. In a haze, the droplets have a radius of up to one micron.

The concentration of hygroscopic nuclei is 10^3 per cubic centimetre over sea, 10^4 per cubic centimetre over land, and 10^5 per cubic centimetre over cities.

Experiments have shown that in the case of scattering by hygroscopic nuclei, that is, in a moist air (but in the absence of a haze) the attenuation is inversely proportional to $\lambda^{1.75}$. In other words, in a moist air the attenuation decreases with increasing wavelength

not so rapidly as in the case of Rayleigh scattering in a dry air. The same experiments [56] have shown that in the visible region of the spectrum the loss due to scattering by hygroscopic nuclei is of the same order of magnitude as that in a dry air. In the presence of a haze, as already noted, the attenuation of visible light and infra-red radiation increases considerably.

Attenuation of Radio Waves by Solids in the Atmosphere. The solid particles (dust, smoke, soot, ash, and the like) that may sometimes find their way into the atmosphere, can behave as nuclei of condensation if they are small in size (not over one micron) and can be wetted by water. When this happens, the attenuation is about the same as in the presence of a haze. The larger particles act as scattering areas and produce a dry haze.

3.15. Tropospheric Propagation of Radio Waves at Optical Frequencies

In contrast to optical sources of visible and infra-red light, lasers (quantum-mechanical devices capable of amplifying or generating coherent light) emit highly monochromatic optical radiation in the form of nearly plane waves which may be treated as having the properties of the radio waves in the more familiar wavelength bands and governed by all laws applicable to conventional radio waves.

A major advantage of using laser waves for communication is that they may be focused into extremely narrow beams emitted by aerials, much narrower than those obtainable in the centimetric band.

As will be recalled, the beamwidth of the main lobe to half-power points is given by

$$\alpha = 1.22\lambda/d \text{ radians} = 70\lambda/d \text{ degrees} \qquad (3.83)$$

where d is the aerial diameter, and λ is the wavelength.

With $d = 10$ centimetres and $\lambda = 1$ micron, the beamwidth will be 2.5 seconds. To obtain the same directivity at a wavelength of three centimetres, one would have to use an aerial three kilometres in diameter. The energy concentrated in the beam is so large that we may safely ignore reflection from the earth's surface and treat the millimetric waves and those in the light wavelength region as ones freely propagated in the troposphere.

However, as they are propagated in the troposphere, the millimetric and optical radio waves experience attenuation and bending due to atmospheric refraction. The bending of the millimetric and laser waves is about the same as that of visible light. That is, they are bent less than ultra-short waves because the polar molecules of

water vapours at so high frequencies have almost no effect on the refractive index of the troposphere.

Owing to the high concentration of energy in the laser beam, even a small receiving aerial (which actually is a system of optical lenses) can extract nearly all energy from the wave. Fig. 3.41 is an idealised sketch of wave propagation in the optical region, where d_1

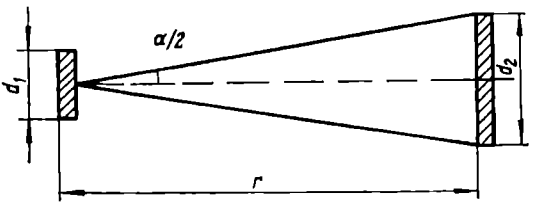


Fig. 3.41. Propagation of radio waves in the optical band

and d_2 are the diameters of the transmitting and receiving aerials, respectively, and r is the propagation path, or the distance between them. As the angle $\alpha/2$ is very small, we may write

$$d_2 = \alpha r \text{ m} \tag{3.84}$$

Taking the numerical values of the previous example, it is not difficult to see that a receiving aerial 10 centimetres in diameter will intercept all the energy contained in the main lobe of the radiation pattern at a distance of 8.2 kilometres, while an aerial with a diameter of one metre will do that at a distance of up to 82 kilometres.

Of course, the wavelength to be used must lie within "transparent windows". It should also be remembered that heavy rain, snow and fog may completely disrupt the radio link operating at laser frequencies. This is why laser links will be practicable only in areas where rain and fog are a rare occurrence. Naturally, all limitations are invalid if the transmitter is above the troposphere and the link or links are with a spacecraft or probe.

Since the main roadblock to using laser links is atmospheric precipitation, it has been suggested [58] to guide the laser beam from transmitter to receiver inside a hermetically sealed pipe. The sole function of the pipe will be to protect the propagation medium from water droplets, and not to serve as a waveguide. Naturally, the segments of the pipe must be perfectly straight. With repeaters located at appropriate intervals, or hops, intelligence can be transmitted by laser links over large distance in much the same way as it is by conventional microwave links.

The amount of information carried by a communication channel is roughly proportional to its frequency. Therefore, it is easy to see the mammoth handling potentialities of laser links with their extremely high frequencies. In this respect, laser links would outdo any other radio circuit used to date.

One more distinction about laser beams is worth while mentioning. From optics it is known that focusing the beam of light generated by usual incoherent sources cannot in principle produce a localized flux density greater than it is at the luminous surface. In contrast, the highly monochromatic coherent radiation of lasers can be focused to a spot whose diameter is comparable with the wavelength. By way of example, if a laser operating at 300 terahertz ($\lambda = 1$ micron) radiates 4 watts, focusing its beam to a spot $4\lambda^2 = 4$ square microns $= 4 \times 10^{-12}$ square metre in area will produce a flux density of $4 \times 10^{12}/4 = 10^{12}$ watts per square metre, which is nearly a million times greater than the flux density at the Sun's surface. It is obvious that this property can be put to many uses. One is to melt the most refractory metals.

A wave normally incident upon an ideally reflecting surface produces a light pressure given by

$$p = 2S/c \ \text{N/m}^2 \tag{3.85}$$

where c is the velocity of wave propagation.

Substituting the found value of flux density in (3.85) gives

$$p = \frac{2 \times 10^{12}}{3 \times 10^8} = 0.7 \times 10^4 \ \text{N/m}^2$$

Noting that 9.81×10^4 newtons per square metre $= 1$ kilogram per square centimetre, we may write

$$p = \frac{0.7 \times 10^4}{9.8 \times 10^4} = 0.071 \ \text{kg/cm}^2$$

For further details concerning laser links, the reader is to refer to [59].

Chapter Four

The Ionosphere

4.1. The Constitution and Structure of the Upper Atmosphere

The term ionosphere applies to that part of the earth's atmosphere which extends out 60 kilometres above the globe.

Since the ionosphere is part of the earth's upper atmosphere, it appears logical to begin our study with the physical properties of the latter, namely its constitution, density (number of molecules per unit volume), and temperature gradient with height.

These properties can be measured either directly or indirectly. Direct methods are concerned with pressure, temperature, air currents and variations in constitution with height and use high-altitude and sound-balloons and radiosondes. The more recent trend has been to use manometers, mass-spectrographs, electron-concentration meters and other instrumentation carried aloft by geophysical rockets, artificial earth satellites and space probes. Sounding balloons and radiosondes are capable of covering altitudes up to forty kilometres. Rockets, satellites and space probes can cut across the entire atmosphere.

Indirect methods include observations of auroral displays (including their spectral analysis), nocturnal light, meteor trails, propagation of radio waves, including those emitted by the Sun, stars and nebulae, atmospherics, and sound measurements.

A very important method for studies of the upper atmosphere is observations of radio propagation by ionospheric stations taking vertical, oblique and oblique-incidence backscatter soundings.

In recent years, the electron concentration of the ionosphere has come to be measured with high-power radar systems which register the scattered radiation due to the electrons present in the ionosphere.

All of these methods, used separately or in combinations, have yielded a wealth of information about the upper atmosphere.

Observations have shown that up to 90 kilometres above the earth the atmosphere has about the same constitution as it has near the surface. It is believed that the ascending, descending and hori-

zontal air currents that exist there mix the air well so that its constitution remains constant despite the difference in mass between the constituent gases.

At greater heights, the atmosphere is stratified owing to this difference in mass, so that the heavier gases are found in the lower part. This part consists mainly of molecular nitrogen and oxygen. The constitution of dry air at sea level is given in Table 4.1.

Table 4.1. CONSTITUTION OF DRY AIR AT SEA LEVEL

Gas	Molecular weight	Partial pressure, mb	Molecules per cubic centimetre	Per cent by volume	Per cent by weight
Dry air	29.0	1012.00	2.7×10^{19}	100.00	100.00
N_2	28.0	791.00	2.11×10^{19}	78.09	75.53
O_2	32.0	212.00	5.6×10^{18}	20.95	23.14
Ar	39.9	9.45	2.5×10^{17}	0.93	1.28
CO_2	44.0	0.31	8.1×10^{15}	0.03	0.046
Ne	20.0	1.2×10^{-2}	4.9×10^{14}	1.8×10^{-3}	1.25×10^{-3}
He	4.0	5.3×10^{-3}	1.4×10^{14}	5.24×10^{-4}	7.24×10^{-5}
H_2	2.0	5.1×10^{-4}	1.35×10^{13}	5×10^{-5}	3.48×10^{-6}

As regards the constitution of air, it is important to remember that in a rarefied atmosphere solar radiation brings about the dissociation of oxygen and nitrogen. That is, the molecules of oxygen and nitrogen, as they absorb a quantum of radiant energy, break down into atoms

$$O_2 + h\nu \rightarrow O + O \text{ and } N_2 + h\nu \rightarrow N + N,$$

where ν is the frequency of the radiation, and $h = 6.62 \times 10^{-34}$ joule-second is Planck's constant.

The energy of a photon causing an oxygen molecule to dissociate corresponds to a wavelength of 0.25 micron, and the energy of a photon required for a nitrogen molecule to dissociate corresponds to a wavelength of 0.168 micron. Since the visible wavelength region of the spectrum extends from 0.4 to 0.75 micron, it is obvious that the dissociation of oxygen and nitrogen can only be caused by ultraviolet or X-rays.

Observations carried out through geophysical rockets have shown that O_2 begins to dissociate at an altitude of about 90 kilometres. At 130 kilometres above the earth, the atmosphere still retains 25 per cent of its O_2. The dissociation of N_2 begins at a height of about 220 kilometres. Besides, the lower atmosphere contains water vapour

to altitudes of about 10 kilometres. The average content of water vapour (by weight) is one per cent, sometimes rising to two or three per cent.

Diagrammatically, the approximate constitution of the atmosphere is shown in Fig. 4.1. The percentage of a particular gas (in terms of molecules or atoms) is laid off as abscissa, and the height as ordinate. As is seen from Fig. 4.1, the region of a constant constitution extends out to 90 kilometres, while above that height the atmosphere begins to stratify.

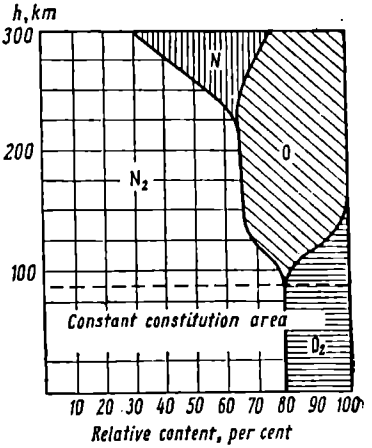


Fig. 4.1. Constitution of the atmosphere (by volume) at different altitudes

Out to 290 kilometres, the dominant constituent of the atmosphere is molecular nitrogen. This is because nitrogen is the main constituent of the atmosphere within the region of constant constitution and the content of molecular nitrogen decreases at a slower rate with height than that of molecular oxygen owing to its smaller molecular weight in comparison with oxygen. Molecular oxygen practically disappears at heights above 160 kilometres. In the interval of heights from 160 to 230 kilometres the atmosphere consists of 65 per cent molecular nitrogen and 35 per cent atomic oxygen. At 300 kilometres and higher, atomic nitrogen becomes the dominant constituent of the atmosphere.

In the interval of heights from 30 to 60 kilometres, solar radiation produces an ozone region (not shown in Fig. 4.1), while between 60 and 90 kilometres, as theoretical studies suggest, chemical reactions lead to the formation of NO. According to calculations, the molecular density of NO at 75-77 kilometres may be 10^8 per cubic centimetre, which works out to about one-millionth of the total number of molecules.

According to the present-day concepts, the terrestrial atmosphere extends out for a distance of two or three earth radii, and is defined as the gaseous envelope that participates in the earth's rotation. The outer atmosphere is composed of ionized particles trapped by the terrestrial magnetic field and travelling along the magnetic lines of force and around the earth. These particles form radiation belts girdling the globe. During geomagnetic disturbances, the ionized gas rotating along with the earth expands to a considerable distance (to about ten earth radii during especially strong magnetic storms).

The most abundant in the outer atmosphere are the atoms and ions of hydrogen and helium, the lightest gases of all.

In the Soviet Union, ionospheric studies have been conducted with the mass spectrometers carried by large geophysical rockets [60] and Sputnik-3. Sputnik-3 has also yielded data on the ion composition of the atmosphere at heights from 225 to 980 kilometres between latitudes 27° and 65° North [61], [62]. From these observations it has been deduced that the predominant ions at those altitudes by day are O^+. Other ions have also been registered, namely those with mass number 14—atomic nitrogen N^+, mass number 18—an isotope of oxygen $(O^{18})^+$, mass number 28—molecular nitrogen N_2^+, mass number 30—nitric oxide $(NO)^+$, and mass number 32—molecular oxygen O_2^+.

The number of molecules in one cubic centimetre of air gives the molecular density of the atmosphere at a particular height h above sea level. It is obvious that the density of air is proportional to the pressure p exerted by the overlying column of air. Therefore, in order to find the height-distribution of molecular density, it will suffice to determine changes in atmospheric pressure with altitude.

As the height above sea level changes by dh, the pressure of air decreases by

$$dp = -\rho g dh \text{ N/m}^2 \tag{4.1}$$

where p = pressure at height h, N/m²;
ρ = density of air, kg/m³;
g = acceleration due to gravity, m/s².

The equation of state, referring to one kilogram-molecule (or kilogram-mole) of a gas, has the form

$$pV = NkT = RT \text{ joules/kg-mole} \tag{4.2}$$

where V = volume occupied by one kilogram-mole, m³/kg-mole;
$N = 6.02 \times 10^{26}$ per kilogram-mole = Avogadro's number, that is the number of molecules per kilogram-mole;
k = Boltzmann's constant = 1.38×10^{-23} joules/°C;
R = universal gas constant = 8.316×10^3 joules/kg-mole °C.

Denoting the mass of a molecule as m, the density of air may be defined as

$$\rho = Nm/V \text{ kg/m}^3 \tag{4.3}$$

Substituting (4.3) in (4.2) gives

$$p(Nm/\rho) = NkT \text{ joules/kg-mole}$$

whence

$$\rho = pNm/NkT = pM/RT \text{ kg/m}^3 \tag{4.4}$$

where $M = Nm$ is the mass of one kilogram-mole.

Substituting ρ from (4.4) in (4.1) yields

$$dp = -\frac{pMg}{RT}dh \text{ N/m}^2$$

which is a differential equation with separable variables

$$dp/p = -dh/H \tag{4.5}$$

where

$$H = RT/Mg \text{ m} \tag{4.6}$$

is the scale height.

Integrating (4.5) gives

$$p = p_0 e^{-h/H} = p_0 e^{(-Mg/RT)h} \text{ N/m}^2 \tag{4.7}$$

where p_0 is the atmospheric pressure in newtons per square metre at the earth's surface. To find the pressure in kg/m^2, its value in newtons per square metre should be divided into 9.81×10^4.

Equation (4.7), known as the barometric equation, relates pressure and height above the earth's surface.

Within the constant-constitution region, M is 29.0, which is the arithmetic-mean weight of the constituent gases of dry air in that region (which extends out to a height of 90 kilometres). At 130 to 250 kilometres, M is 25, and at greater altitudes, it is 15.

Physically, the scale height is the height of a constant-density atmosphere that would produce the same pressure as the actual one. At $T = 273°$K, it is eight kilometres for the constant-constitution (or completely mixed) atmosphere, and 7.2 kilometres for molecular oxygen.

The same equation (4.7) can be used to determine the partial pressure of atmospheric gases in the region where they are stratified according to their molecular weights, with the difference that instead of p_0, or the pressure at sea level, use should be made of p_{h0}, the pressure at the height where the stratification begins. The total pressure at a given height is equal to the sum of the partial pressures.

There is an additional difficulty in determining the relationship between pressure and height since the temperature of air likewise changes with altitude. Because of this, in determining the pressure in the constant-constitution region use is made of an integral equation of the form

$$p_h = p_0 e^{-\frac{Mg}{R}\int_0^h \frac{dh}{T(h)}} \text{ N/m}^2 \tag{4.8}$$

where $T(h)$ is the temperature as a function of height.

Recent studies into the height distribution of pressure with rockets, satellites and spacecraft and also the values of the aerodynamic drag they experience upon re-entry have shown that the earth's atmosphere is much denser than it was expected to be. It has also been found that the density distribution with height, especially above 100-200 kilometres, is subject to variations due to changes in the thermal state of the atmosphere (notably, due to solar flares). Therefore, we may only speak of average values of density. At present, the molecular density is assumed [63] to be 3×10^{13}, 2×10^{10} and 10^9 per cubic centimetre at 100, 200, and 300 kilometres, respectively.

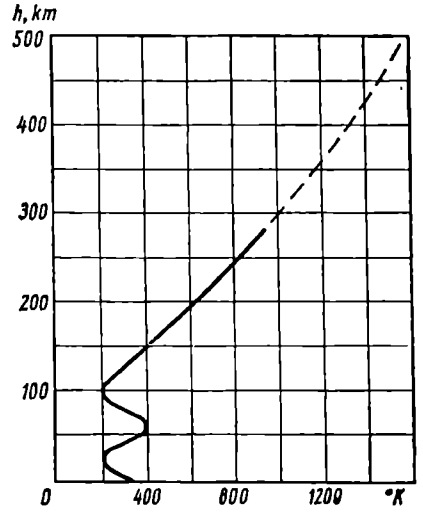


Fig. 4.2. Air temperature versus height

The temperature distribution with height in the troposphere was discussed in detail in § 3.1. At the upper boundary of the troposphere, the temperature ceases to drop and levels off at 210°K (—60°C) around 20 kilometres. At greater altitudes, the temperature begins to rise, reaching a maximum of approximately 400°K at 60 kilometres or so. At about 80 kilometres, a temperature minimum (200°-250°K) occurs, after which the temperature rises to a thousand degrees or more. Like the molecular density, at altitudes in excess of 200 kilometres the temperature of the atmosphere is subject to marked variations during solar flares and the influx of charged particles into the terrestrial atmosphere. The average distribution of temperature up to 500 kilometres is shown in Fig. 4.2.

The temperature peak at 60 kilometres is due to the formation of O_3. A second rise in temperature at greater altitudes is caused by the absorption of solar radiation as the ionized region is formed.

The temperature of a rarefied atmosphere should be construed in terms of the kinetic theory of gases, namely, as a quantity describing the average velocity of random motion of molecules, atoms and electrons. Exactly this meaning is put into the concept of the electron temperature of the ionosphere. For the same reason, the temperature of the upper atmosphere is often called kinetic.

It should be remembered that the upper atmosphere, like the troposphere, is influenced by winds and vertical air currents which produce turbulences and local irregularities. Because of this, the upper atmosphere is a time-variant formation subject to fluctuations.

4.2. The Mechanism of Ionization

The atoms of atmospheric gases consist of charged nuclei and electron shells around them. In neutral atoms, the positive charge of the nucleus is balanced by the negative charge of the peripheral electrons.

When one (seldom, several) electron is broken loose from the outer shell of the atom, the electric balance is upset, and the atom acquires one (or several) elementary charge. This is ionization. Since the electrons on the outer shell are attracted to the positive nucleus, some energy must be expended to break them loose. The minimum energy required to remove an electron from its shell and send it into field-free space is called the *work function*. The ionization work function has been accurately measured for all atmospheric gases, both in the molecular and atomic states, under laboratory conditions.

Ionization can occur by several distinct mechanisms. We shall limit ourselves to *ionization by absorption of radiation* (or *photo-ionization*) and *ionization by collision*.

Writing the ionization work function as W, we may state that when a given gas is exposed to a radiation with the energy of photons $h\nu$, ionization will occur if

$$h\nu > W \text{ joules} \tag{4.9}$$

or, re-writing

$$\nu > \nu_{ion} \tag{4.9a}$$

where

$$\nu_{ion} = W/h \text{ Hz} \tag{4.10}$$

is the frequency of the radiation.

From (4.9a) it follows that a given gas absorbing a radiation will be ionized if the radiation frequency is in excess of a certain critical value, known as the *ionization frequency* (or, in terms of wavelength, the *ionization wavelength*). No increase in the energy of photons at other frequencies will ionize the gas.

When the knocked-out electrons have a low velocity so that the relativistic factors may be neglected, the total energy balance may by written as

$$h\nu = W + mu^2/2 \text{ joules} \tag{4.11}$$

From (4.11) it is seen that with the same number of photons an increase in the frequency of radiation brings about an increase in the velocity of the knocked-out electrons. This is how the things are with ionization by absorption of radiation.

In ionization by collision, the atoms or molecules of a gas lose a peripheral electron as they collide with other particles (corpuscles) having a sufficient kinetic energy. Assuming as before that the relativistic correction for the mass of moving particles may be neglected, the condition of ionization by collision may be written as

$$\frac{m_1 u_1^2}{2} > W \text{ joules} \tag{4.12}$$

or as

$$u_1 > u_{ion} \tag{4.12a}$$

where

$$u_{ion} = \sqrt{2W/m_1} \text{ m/s} \tag{4.13}$$

is the critical velocity of a particle at which ionization occurs.

The total energy balance now is

$$\frac{m_1 u_1^2}{2} = W + \frac{mu^2}{2} \text{ joules} \tag{4.14}$$

where m_1 and u_1 are the mass and velocity of the ionizing particle, and m and u are the mass and velocity of the knocked-out electron.

The energy of a moving particle is customarily expressed in *electron-volts* (eV), assuming that the particle has the charge and mass of an electron. Then the kinetic energy and work function can be related thus

$$\frac{mu^2}{2} = eV \text{ joules} \tag{4.15}$$

where e is the charge of an electron, and V is the potential difference in volts on passing through which an electron originally at rest will acquire the velocity u.

Equation (4.15) uniquely relates the velocity u and the energy of a moving electron expressed in electron-volts. As a proof, we note that one electron-volt is equal to 1.6×10^{-19} joule. Then we may re-write Eq. (4.15) as

$$u = \frac{\sqrt{2\,[eV]_j}}{m_{kg}} = 5.65 \times 10^{-10} \sqrt{\frac{[eV]_{ev}}{m_{kg}}} \text{ m/s} \tag{4.15a}$$

If the product $[eV]$ is numerically equal to the ionization work function W, Eq. (4.15a) reduces to

$$u = 5.65 \times 10^{-10} \sqrt{\frac{W_{ev}}{m_{kg}}} \text{ m/s} \tag{4.15b}$$

To relate the ionization wavelength λ_{ion} and the ionization work function W, let us re-write Eq. (4.10) as

$$\nu_{ion} = W/h = c/\lambda_{ion} \text{ Hz}$$

whence

$$\lambda_{ion} = ch/W \text{ m} \tag{4.16}$$

where W is in joules, and $c = 3 \times 10^8$ metres per second. Since one electron-volt is equal to 1.6×10^{-19} joule, Eq. (4.16) may be re-written as

$$\lambda_{ion} = \frac{ch10^{19}}{1.6W_{ev}} \text{ m} \tag{4.16a}$$

Substituting the numerical values of c and h in Système Internationale units gives

$$\lambda_{ion} = \frac{12.394 \times 10^{-7}}{W_{ev}} \text{ m} \tag{4.16b}$$

or in the more convenient units

$$\lambda_{ion} = \frac{1.2394}{W_{ev}} \ \mu \tag{4.16c}$$

and

$$\lambda_{ion} = \frac{12394}{W_{ev}} \ \text{Å} \tag{4.16d}$$

one Ångstrom unit being equal to 10^{-10} metre.

For ionization by absorption of radiation, it is convenient to express the ionization work function in terms of the ionization wavelength. For ionization by collision, the more convenient form is the ionization work function expressed in electron-volts.

The values of the ionization work functions for the atmospheric gases are summarized in Table 4.2, where the letter e in the reaction equations stands for the electron being released.

Table 4.2. IONIZATION WORK FUNCTIONS OF ATMOSPHERIC GASES

Constituent gas	Reaction	Ionization wavelength, Å	Ionization work function, eV
O_2	$O_2 + h\nu \longrightarrow O_2^+ + e$	1015	12.2
O	$O + h\nu \longrightarrow O^+ + e$	910	13.61
O^+	$O^+ + h\nu \longrightarrow O^{++} + e$	350	35.1
N_2	$N_2 + h\nu \longrightarrow N_2^+ + e$	795	15.51
N	$N + h\nu \longrightarrow N^+ + e$	850	14.53
He	$He + h\nu \longrightarrow He^+ + e$	503	24.58
H_2	$H_2 + h\nu \longrightarrow H_2^+ + e$	802	15.42
H	$H + h\nu \longrightarrow H^+ + e$	910	13.60
NO	$NO + h\nu \longrightarrow (NO)^+ + e$	1340	9.25

As is seen, even the most easily ionizable gas, nitric oxide, can be ionized by the short ultra-violet rays only. It may be noted in passing that only the longest ultra-violet rays with wavelengths in excess of 2900 Ångstroms reach the earth's surface. The shorter ultra-violet rays are completely absorbed by the upper atmosphere, and their energy is taken up in ionization, dissociation, formation of ozone, photo-chemical reactions, etc.

From reference to Table 4.2, it is an easy matter to determine the velocity that a particle with the mass of an electron should have in order to ionize any of the atmospheric gases.

Example 4.1. Determine the velocity u that a moving electron should have in order to ionize molecular oxygen by collision.

Solution: Substituting the mass of an electron, $m = 9.106 \times \times 10^{-31}$ kg in Eq. (4.15b) and the ionization work function for molecular oxygen from Table 4.2, we get

$$u = 5 \cdot 65 \times 10^{-10} \sqrt{\frac{12.2 \times 10^{31}}{9.106}} \cong 2.1 \times 10^{6} \text{ m/s} = 2100 \text{ km/s}$$

Thus, oxygen molecules can be ionized by particles with the mass of an electron only if their velocity is greater than 2100 kilometres per second.

4.3. Ionizing Agencies

The main ionizing agency in the atmosphere is the Sun whose photosphere has a temperature of about 6000° K and emits electromagnetic waves with a broad gamut of frequencies, while its chromosphere and corona heated to 6×10^{5} °K (some believe the temperature of the solar corona to be as high as 2×10^{6} °K) are sources of soft X-rays in the wavelength range from 8 to 20 Ångstroms and of ultra-violet rays between wavelengths of 20 to 300 Ångstroms. Besides, the sun continually releases streams of electrons and other charged particles that produce corpuscular radiation.

From Table 4.2 it follows that ionization can be produced by electromagnetic waves shorter than 1340 Ångstroms, while for the major constituent gases the wavelengths must be shorter than 1000 Ångstroms. These wavelengths fall in the region of ultra-violet (from 20 to 4000 Ångstroms) and soft X-rays.

The intensity of ultra-violet solar radiation can be found by Planck's law, treating the sun as an absolute black body with a temperature of 6000°K. Calculations show that the ultra-violet rays capable of ionization (more specifically, those with wavelengths from zero to 1000 Ångstroms) account only for 9.6×10^{-8} of the

total electromagnetic energy emitted by the sun. Of course, it is impossible to measure this radiation at the earth's surface, because all waves shorter than 2900 Ångstroms are absorbed by the atmosphere. The corpuscular streams fail to reach the earth's surface for the same reason. It is only with the advent of rockets and spacecraft carrying photon and particle counters that the radiations causing ionization in the terrestrial atmosphere have come to be measured directly in recent years [64]. Interesting measurements of corpuscular streams were taken by means of fluorescent screens on *Sputnik-3*.

From the data gathered to date, it appears that the corpuscular radiation of the sun accounts for not over 50 per cent of the sun's total ionizing electromagnetic radiation.

Speaking of the sun as the main source of ionization in the terrestrial atmosphere, it should be remembered that ultra-violet solar radiation in the wavelength region shorter than 2000 Ångstroms does not remain fixed from year to year—it follows the sunspot (eleven-year) cycle in solar activity. There is an unmistakable relationship between the relative sunspot number and the intensity of ionizing radiation. As the number of sunspots increases, the corpuscular radiation of the sun also gains in strength.

From time to time there occurs a considerable increase in the intensity of the shorter ultra-violet rays, which are especially ionizing. It has been noted that this increase always takes place at the time of solar flares which usually last from a few minutes to several hours.

The intensity of corpuscular streams also increases at the time of geomagnetic disturbances. The rate and strength of solar flares and also the rate and strength of geomagnetic disturbances likewise go up as solar activity increases.

A more detailed discussion of the effects that the sunspot (eleven-year) cycle of solar activity has on the state of the ionosphere and of some other irregular processes in the ionosphere will be found in Chapter 5.

The next significant ionizing agency is the effect of the stars. Some of the stars have a temperature of 20,000°K and higher, which goes to build up ultra-violet radiation and compensate to a certain degree for the fact that the stars are far away from the earth. The ionizing effect of stellar radiation is about one-thousandth of that of solar radiation.

At one time it was widely believed that a good proportion of the ionization in the earth's atmosphere was due to cosmic rays because of their extremely high energy (averaging 10^{10} electron-volts and reaching peaks of 10^{19} electron-volts). Direct measurements of the ionizing effect of cosmic rays have shown that this is so in the lower

atmosphere, and cosmic rays have found to produce three or four ions in a cubic centimetre every second there. In the ionosphere, the ionizing effect of cosmic rays is drastically reduced and amounts to just one-hundred millionth of that of ultra-violet solar radiation.

Last but not least, there is cosmic dust that may act as an ionizing agency. Continually impinging upon the earth's atmosphere, these tiny specks of matter burn up in the air and leave areas of ionization behind. These particles are assumed to have a radius of 0.4 to 4 microns, and there are presumably about 10^{20} such particles hitting the atmosphere over a day. The ionizing effect of cosmic dust cannot yet be evaluated quantitatively. The bigger meteors produce ionized trails having the shape of narrow shafts which start out with a diameter of a few centimetres and gradually expand towards the earth's surface. These ionized trails may be several tens of kilometres in length, with a linear electron density of 10^{10} to 10^{16} per metre and a life time of a few tenths of a second to several minutes.

It appears relevant to quote some of the data gathered with rocket-borne photon counters [65] about the sun's ultra-violet and soft X-rays.

The solar constant of radiation, that is, the intensity of solar radiation in free space at the earth's mean solar distance, is believed to be

$$C = 2.00 \text{ gcal per square centimetre minute}$$
$$= 1.39 \times 10^3 \text{ watts per square metre}$$

Measurements have shown that the intensity of solar radiation with a wavelength of 1215.7 Ångstroms (Lyman-alpha radiation) is anywhere from 10^{-4} to 4×10^{-4} watts per square metre. The intensity of solar radiation with a wavelength of 1025.7 Ångstroms (Lyman-beta radiation) is 3×10^{-5} watts per square metre. The overall intensity of soft X- and ultra-violet rays in the wavelength range between 8 and 100 Ångstroms is 10^{-4} watts per square metre. Unfortunately, at the time of writing, data on the region between 100 and 1000 Ångstroms were lacking.

4.4. Ionization in a Homogeneous Atmosphere

For better insight into the mechanism of ionization in the atmosphere, we shall consider the simplest case—a homogeneous atmosphere with the same temperature throughout and with the pressure obeying the barometric equation, on which a monochromatic ionizing radiation of frequency ν is made to impinge. It is assumed that the energy of a photon, $h\nu$, of the ionizing radiation exceeds the ionization work function W. To begin with, let us consider the

case of ionizing radiation at normal incidence and let the flux density of this radiation at the outer edge of the atmosphere be S_∞ (Poynting's vector).

It is obvious that ionization cannot be a maximum either at the earth's surface or at the outer edge of the atmosphere. At the earth's surface this cannot happen because ionizing radiation passes through the entire body of the atmosphere before it reaches the ground, and it is highly unlikely that the radiation can retain enough energy to produce a peak ionization. At the outer edge of the atmosphere, although ionizing radiation is at its strongest there, this cannot happen because the density of the atmosphere is insignificant to a point where the total ionization would be negligible, even if all molecules in the unit volume were ionized. This implies that an ionization peak should occur at some finite height.

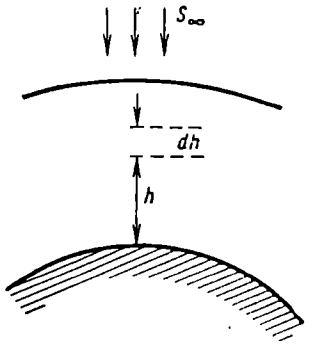


Fig. 4.3. Explaining formation of a maximum ionization zone

Referring to Fig. 4.3 which shows ionization at normal incidence, S is the flux density of ionizing radiation (in watts per square metre) at a height h above the earth's surface, and I_S is the number of free electrons produced by this radiation in unit volume every second.

The flux density of ionizing radiation falls off at the rate dS/dh. Assuming that this rate remains constant over an interval of one metre, dS/dh may be interpreted as the change in flux density with immersion to one metre. Since the flux density is referred to one square metre of cross-sectional area, it follows that the quantity dS/dh, which has the dimensions of watts per second-cubic metre, expresses the loss of energy by ionizing radiation in one cubic metre of the atmosphere at a height h above the earth's surface. Noting that a photon can knock out only one electron and assuming that all energy expressed by dS/dh is taken up in ionization, the number of electrons produced in a cubic metre every second will be given by

$$I_S = \frac{1}{[h\nu]} \frac{dS}{dh} \text{ m}^{-3}\text{s}^{-1} * \qquad (4.17)$$

As it passes through a layer of the atmosphere dh thick, the radiation decreases in flux density by dS, which is directly proportional to the flux density S, the thickness of the elementary layer dh, and

* In this and other equations of this section, the energy of a photon, $[h\nu]$, is put in square brackets to make Planck's constant look different from the height h.

the number of molecules per unit volume, which is in turn proportional to atmospheric pressure. Writing the proportionality factor as B, we may write

$$dS = BSpdh \text{ W/m}^2 \tag{4.18}$$

Describing pressure as a function of height with the barometric equation and setting $Mg/RT = b$ (metre^{-1}) and $Bp_0 = B'$ (metre^{-1}) we may re-write Eq. (4.18) as follows

$$dS = B'Se^{-bh}dh \text{ W/m}^2 \tag{4.19}$$

This is a differential equation with separable variables, whose solution gives the flux density of ionizing radiation as a function of height.

Separating the variables and integrating the left-hand side from S to S_∞, and the right-hand side from h to ∞, that is over the path that ionizing radiation covers to reach a point at height h,

$$\int_S^{S_\infty} \frac{dS}{S} = B' \int_h^\infty e^{-bh}\, dh \tag{4.20}$$

we get

$$\ln \frac{S_\infty}{S} = B'e^{-bh} \tag{4.21}$$

Here it is assumed that the outer edge of the atmosphere is at an infinitely large, actually sufficiently large, distance from the earth's surface.

Solving the logarithmic equation (4.21) for S gives

$$S = S_\infty e^{-\frac{B'}{b} e^{-bh}} \text{ W/m}^2 \tag{4.22}$$

Substituting (4.22) in (4.19) and the latter in (4.17) gives

$$I_S = \frac{B'S_\infty}{[h\nu]} e^{-\left(\frac{B'}{b}e^{-bh} + bh\right)} \text{ m}^{-3}\text{s}^{-1} \tag{4.23}$$

Equation (4.23) shows how the number of free electrons produced in unit volume every second varies with height.

Let us find the height h_0 at which I_S is a maximum. Noting that h enters only the exponent, differentiating the latter with respect to h and equating the result to zero gives

$$-B'e^{-bh} + b = 0$$

whence

$$h_0 = \frac{1}{b} \ln \frac{B'}{b} \text{ m} \tag{4.24}$$

Equation (4.24) gives substance to the intuitive assumption that ionization should be a maximum at a finite height. As is seen, this height is directly proportional to the atmospheric absorption B' and the temperature of the atmosphere, and inversely proportional to the average molecular weight of the gases being ionized.

Equation (4.23) can be readily extended to the more general case of ionizing radiation at oblique incidence, when the zenith angle

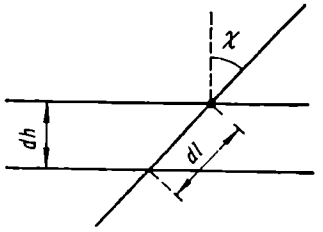


Fig. 4.4. Determining an element of path for ionizing radiation at oblique incidence

of the sun (the angle χ between the direction of sun rays and the vertical) is non-zero.

On passing through dh, ionizing radiation covers a path (Fig. 4.4)

$$dl = dh/\cos\chi \text{ m} \tag{4.25}$$

In the process, it produces as many free electrons as is given by

$$I_S = \frac{1}{[h\nu]} \frac{dS}{dl} \text{ m}^{-3}\text{s}^{-1} \tag{4.26}$$

Substituting (4.25), we may write (4.26) as follows

$$I_S = \frac{1}{[h\nu]} \frac{dS}{dh} \cos\chi \text{ m}^{-3}\text{s}^{-1} \tag{4.26a}$$

Doing the same in (4.18) and (4.19), we get

$$dS = BSp\,dl = B'Se^{-bh} \frac{dh}{\cos\chi} \text{ W/m}^2 \tag{4.27}$$

Now, using what we have found for normal incidence, instead of (4.23) we get

$$I_S = \frac{B'}{[h\nu]} S_\infty e^{-\left(\frac{B'}{b\cos\chi}e^{-bh}+bh\right)} \text{ m}^{-3}\text{s}^{-1} \tag{4.28}$$

Ionization will then be a maximum at a height given by

$$h_0 = \frac{1}{b} \ln\left(\frac{B'}{b\cos\chi}\right) \text{ m} \tag{4.29}$$

Equation (4.28) relates the number of free electrons produced in every cubic metre every second to height and zenith angle, and Equation (4.20) describes the height of a peak ionization as a function of zenith angle.

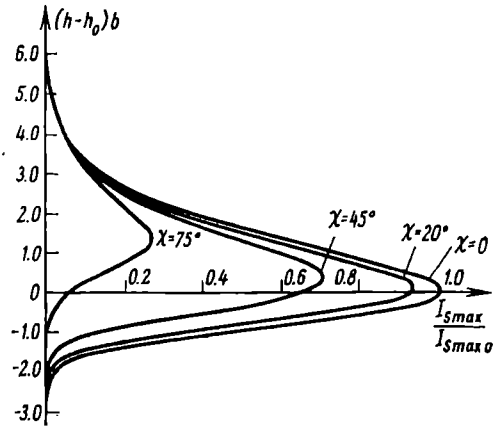


Fig. 4.5. Relative number of free electrons produced in one cubic metre of air in one second as a function of height and zenith angle of the sun

Substituting (4.29) in (4.28), we find an expression for the maximum value of I_S

$$I_{S\,max} = \frac{S_\infty b}{[h\nu]\,e} \cos\chi \ \text{m}^{-3}\text{s}^{-1} \tag{4.30}$$

Writing the maximum number of electrons freed in every cubic metre every second at zero zenith angle as $I_{S\,max\,0}$, we may re-write (4.30) as

$$\frac{I_{S\,max}}{I_{S\,max\,0}} = \cos\chi \tag{4.31}$$

A plot of the ratio $I_{Smax}/I_{Smax\,0}$ as a function of the height h above the earth's surface for four values of zenith angle appears in Fig. 4.5. The relative height (the one relative to h_0 at which ionization is a maximum when $\chi = 0$) is laid off as ordinate. The distances are laid off to a scale given by $1/b = H = RT/Mg$, which is the scale height. As is seen from the plot of Fig. 4.5, the height of the maximum value of I_S increases with increasing zenith angle.

From Equations (4.23) and (4.28) it follows that with assumptions made, that is, a homogeneous atmosphere, constant temperature and monochromatic radiation, there will appear an area of ionization with a single peak, as shown in Fig. 4.5. This is often referred

to as a simple ionized layer. In the Soviet Union, its theory has been formulated by S. I. Kryuchkov [66], and abroad by S. Chapman [67], and the layer itself is often called the Chapman layer.

4.5. Disappearance of Free Charges in the Atmosphere

So far, we have treated ionization as a one-way process. It has been assumed that ionizing radiation of one kind or another produces I_S free electrons per unit volume of air per second and, obviously, as many positive ions. Treating the process of ionization in this way, it is not difficult to see that if unit volume initially contains N_m neutral molecules all the molecules will be completely ionized in $\tau = N_m/I_S$ seconds.

Actually, there is more than that to the process of ionization. For one thing, ionization is accompanied by recombination—free electrons re-combine with positive ions. Given certain conditions, the two processes—ionization and recombination—may strike a kind of dynamic balance: as more free electrons are produced, an equal number of them recombine with positive ions.

The physical cause of recombination lies in the random thermal motion of electrons, ions and neutral molecules, owing to which electrons and ions happen to move so closely to one another that they merge together under the action of electrostatic attraction producing a neutral atom or a neutral molecule.

What happens may be described as follows

$$\left.\begin{aligned} O^+ + e &\rightarrow O + h\nu \\ O_2^+ + e &\rightarrow O_2 + h\nu \end{aligned}\right\} \tag{4.32}$$

where O^+ stands for a positive ion. The recombination of other atmospheric gases may be described in a similar manner.

As follows from (4.32), the recombination of an electron and an ion is accompanied by emission of energy in the form of a photon, equal to the ionization work function. Thus, the energy spent to ionize a gas is partly liberated in recombination. We have said "partly", because the excess energy is converted into the kinetic energy of the electrons broken loose.

Let there be one electron and one positive ion in one cubic metre of air. The probability that they will recombine is α_e, which is called the *electron-ion recombination coefficient*. Its physical meaning is that during a time interval equal to $1/\alpha_e$ a given electron will collide with a positive ion once. In the adopted units, the electron-ion recombination coefficient has the dimensions of cubic metre per second.

Now let there be one more electron and one more positive ion added to the same unit volume of air (Fig. 4.6). Now the probability

of recombination increases four-fold, because there is an equal probability of recombination for the previous electron and the previous ion, the previous electron and the new ion, the new electron and the previous ion, and the new electron and the new ion.

Applying the same reasoning to N electrons, we may write

Number of electrons	Probability of recombination
1	α_e
2	$4\alpha_e$
3	$9\alpha_e$
.	.
.	.
.	.
N	$N^2\alpha_e$

The number of free electrons, N, per unit volume is known as *electron concentration* or *electron number density*.

If ionizing radiation produces I_S free electrons per unit volume of air per second and if recombination removes $\alpha_e N^2$ electrons, the equation of state for an ionized gas may be given the form

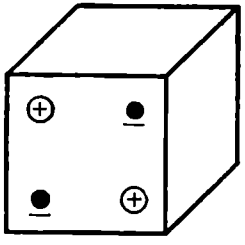


Fig. 4.6. Recombination of electrons

$$\frac{dN}{dt} = I_S - \alpha_e N^2 \text{ m}^{-3}\text{s}^{-1} \qquad (4.33)$$

known as the *continuity equation*.

Consider two cases.

(a) *Dynamic Balance.* When the number of electrons produced is equal to those disappearing due to recombination, then $dN/dt = 0$ and (4.33) reduces to

$$I_S - \alpha_e N^2 = 0$$

so that

$$N = \sqrt{I_S/\alpha_e} \text{ m}^{-3} \qquad (4.34)$$

From (4.34) it follows that if the recombination coefficient is assumed to vary little with height, the height at which N reaches its maximum values occurs near the height at which I_S is a maximum. Consequently, the conclusion we drew in the previous section about the height of the maximum value of I_S may be extended to include N, the electron number density.

It can also be seen that the absolute value of N is inversely proportional to the square root of the recombination coefficient. That is, as the rate of recombination slows down, the electron number density reaches a greater value in conditions of dynamic balance.

(b) *Recombination upon Sudden Interruption of Ionizing Radiation.* This may, for example, happen after sun-set. In this case, the time

dependence of N can likewise be determined from Eq. (4.33). Let at time $t = 0$, that is, when ionizing radiation ceases, the electron number density be N_0. Then Eq. (4.33) reduces to

$$\frac{dN}{dt} = -\alpha_e N^2 \text{ m}^{-3}\text{s}^{-1} \tag{4.35}$$

Separating the variables and integrating the left-hand side from N_0 to N and the right-hand side, from 0 to t,

$$\int_{N_0}^{N} \frac{dN}{N^2} = -\alpha_e \int_0^t dt \tag{4.36}$$

we get

$$N = \frac{N_0}{1 + N_0 \alpha_e t} \text{ m}^{-3} \tag{4.37}$$

Equation (4.36) shows that as t increases, the electron number density decreases hyperbolically, tending to zero in the limit. The rate of decrease of electron number density [see (4.35)] is decided by the recombination coefficient—the greater the absolute value of α_e, the faster the rate of decrease. Eq. (4.36) explains why at nightfall, that is, after ionizing radiation ceases, the electron number density decreases more or less gradually, and not suddenly, so that throughout a winter night there remains a sufficient number of electrons per unit volume to reflect radio waves.

What we have just discussed is known as simple recombination. It can be observed in the upper atmosphere where neutral molecules are few and far between. In the denser layers of the atmosphere, where in addition to electrons and ions there is a large number of neutral molecules, that is, where only a small fraction of the total molecules is ionized, multi-step recombination takes place. In sketch form, multi-step recombination is shown in Fig. 4.7. Owing to thermal motion, free electrons may move very closely to molecules and attach themselves to them. This is extremely likely to occur because there is a great number of molecules present and there exists a force of attraction between the electrons and the neutral molecules (described in terms of the attachment coefficient). This is the first step in recombination. Then, the negative ion thus formed combines with a positive ion to produce two neutral molecules. Recombination as a whole is described in terms of the effective recombination coefficient, α_{ef}. In real gases, recombination is still more complicated [68].

The complete equation of state for an ionized gas should take into account the fact that free electrons are produced not only through direct ionization (I_S), but also through photo-detachment (that is, removal of electrons from negative ions by photons). On the other

hand, free charges may disappear owing to (1) the recombination of electrons and positive ions, (2) the attachment of electrons to neutral molecules, and (3) recombination of negative and positive ions.

A very important characteristic of an ionized gas, or plasma, is the collision frequency, ν, or the average number of collisions per second that an electron experiences with neutral molecules, ions, and electrons. These collisions ought not to be identified with those resulting in recombinations. Moving about at random due to thermal agitation, an electron collides with these particles to impart

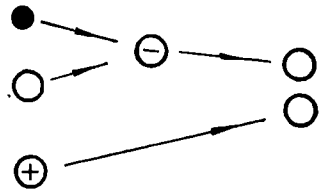


Fig. 4.7. Multi-step recombination

them some of its kinetic energy. But only a small fraction of the collisions with ions results in recombinations and only in some of the collisions with neutral molecules do electrons attach themselves to the molecules.

4.6. Ionization in the Real Atmosphere

In the real atmosphere ionization differs from that occurring in a homogeneous atmosphere having a constant temperature and acted upon by a monochromatic radiation in at least three aspects.

To begin with, the real atmosphere is not homogeneous. It is assumed to be completely mixed only to a height of about 90 kilometres. Above that height, it is apparently stratified.

Nor is the temperature of the atmosphere constant with height. The expected distribution of temperature with height is shown in Fig. 4.2.

Finally, ionization in the atmosphere is produced not only by solar ultra-violet and X-rays (with a broad range of frequencies) but also by corpuscular streams.

Taken together, these factors make the electron-density distribution differ markedly from the one discussed in § 4.4.

At the time when information about the structure of the ionosphere was supplied solely by ground-based ionospheric stations and the electron number density was determined from the measured values of frequency at which radio waves at vertical incidence would

still be reflected from the ionosphere, the opinion was widely held that the atmosphere had at least three well-defined regions, denoted *D*, *E*, and *F* (in order of increasing height), and the respective layers denoted *D*, *E1*, *E2*, *Es* (or sporadic *E*), *F1*, and *F2*. For each of the *E* and *F* layers, the ionospheric stations had determined so-called critical frequencies, that is, the highest frequencies that could be reflected from regions of maximum electron number density.

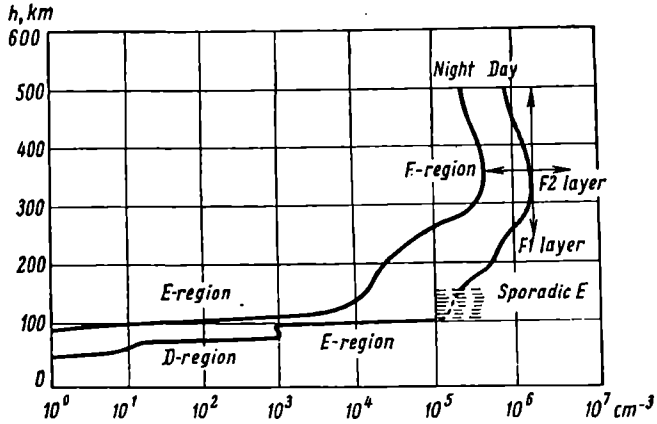


Fig. 4.8. Approximate distribution of electron density with height by day and at night. The arrows show the limits of variations for the height and electron density for the *F2* layer

Explorations of the ionosphere with rockets and artificial satellites, notably using a dispersion interferometer [65], [69], have shown that the well-defined regions or layers in the ionosphere are non-existent. One only may speak of one well-defined peak—in the *F2* layer and three more just discernible peaks or, rather, areas with insignificant variation in electron density.

In the daytime, the *D* region extends between heights of 60 and 90 kilometres; the *E* region, between 100 and 140 kilometres; the *F1* layer, between 180 and 240 kilometres; and *F2* between 230 and 400 kilometres. At night, the *D* region and the *F1* layer disappear, but the *E* region and the *F2* layer remain. Naturally, their electron density materially decreases at night. An approximate distribution of electron density among the various regions and layers of the atmosphere by day (when the atmosphere is exposed to solar radiations) and at night is shown in Fig. 4.8 after [70].

Fig. 4.9 shows a vertical profile of electron density as determined by two methods with geophysical rockets [69]. The key characteri-

stics of the various regions and layers of the atmosphere are summarized in Table 4.3.

In addition to the data presented in Table 4.3, it is important to note the following. The *D* and *E* regions and also the *F1* layer are highly stable—diurnal variations in electron density and the

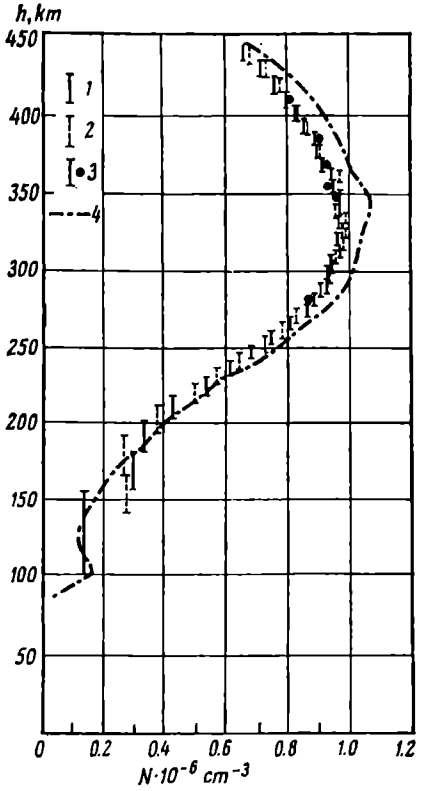


Fig. 4.9. Vertical electron density profile (as plotted by means of a Soviet geophysical rocket August 27, 1958)

heights of the layers remain about the same from day to day. In the *E* region and the *F1* layer, the electron number density in the daytime is uniquely decided by the zenith angle χ and is given by

$$N = N_{max\,0} \sqrt{\cos \chi} \text{ m}^{-3} \tag{4.38}$$

where N_{max0} is the maximum value of N in a given region or layer when $\chi = 0$.

Table 4.3. CHARACTERISTICS OF THE VARIOUS REGIONS IN THE IONOSPHERE

Particulars	*D* region	*E* region	*F1* layer	*F2* layer
Likely origin	(a) Ionization of NO with Lyman-alpha radiation (b) Ionization of all gases by soft X-rays	Ionization of all gases by soft X-rays	Ionization of O with fast decrease of recombination coefficient with height	Ionization of O by UV, X-rays and probably corpuscular radiation
Height, km	60-90 by day; disappears at night	100-140	180-240 by day; disappears at night	230-400
Molecular density, per cubic centimetre	10^{14}-10^{16}	5×10^{11}-10^{13}	about 10^{11}	about 10^{10}
Electron or ion density, per cubic centimetre	100-10^3 for electrons; 10^6-10^8 for ions	up to 10^5 to 4.5×10^5 by day; fixed at about 5×10^3 to 10^4 at night	2×10^5-4.5×10^5	max. 2×10^6 by day in winter; max. 2×10^5 by day in summer, 3×10^5 at night in winter
Collisions, per second	10^7 at lower edge	10^5	10^4	10^3-10^4
Recombination coefficient, cu. cm per sec	10^{-5}-10^{-7}	10^{-8}	4×10^{-9}	8×10^{-11} by day; 3×10^{-11} at night

After nightfall, the *D* region disappears because of the high rate of multi-step recombination, and so does the peak characterizing the *F1* layer (see Fig. 4.8). For the reasons not yet elucidated, the electron density of the *E* region remains fixed at about 5×10^3 to 10^4 electrons per cubic metre throughout the night.

The *F2* layer exhibits a number of conspicuous irregularities. The vertical and horizontal arrows in Fig. 4.8 show the ranges of variations in the height and electron density in the *F2* layer with geogra-

phical location, solar activity, and local time. As is seen, electron density and the height of the layer vary from day to day considerably. Various disturbances likewise occur very often here. The critical frequency has two maximum values—one for winter, and the other for summer (Figs. 4.10 and 4.11). The relationship between the electron density and the critical frequency is given by

$$N = 1.24 \times 10^4 f^2_{c(\text{MHz})} \text{ cm}^{-3} \tag{4.39}$$

For the sake of comparison, these plots also include the respective curves for the E region and the $F1$ layer.

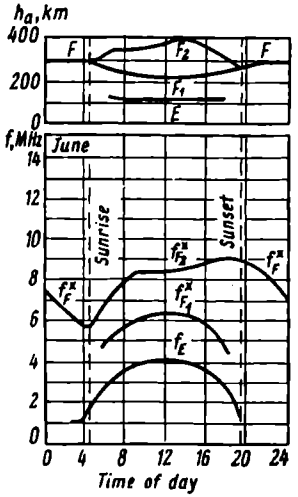


Fig. 4.10. Averaged daily plots of critical frequencies and virtual heights of the E and F regions in summer

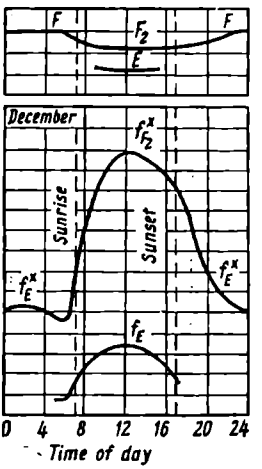


Fig. 4.11. Averaged daily plots of critical frequencies and virtual heights of the E and F regions in winter

Furthermore, the critical frequency reaches its maximum well after noon, which does not happen in the E region and the $F1$ layer. The winter N curve for the $F2$ layer has a higher maximum which occurs a little after local noon and a deep minimum occurring just before dawn. The summer curve is more quiet, which is apparently explained by the heating and convection of the air masses. The intermediate forms are not well-defined, and either the winter or the summer form predominates in spring and autumn. A change-over from the winter to the summer characteristics and the irregularities of the $F2$ layer, as observed in Japan, are illustrated in Fig. 4.12. The dots represent the individual measurements.

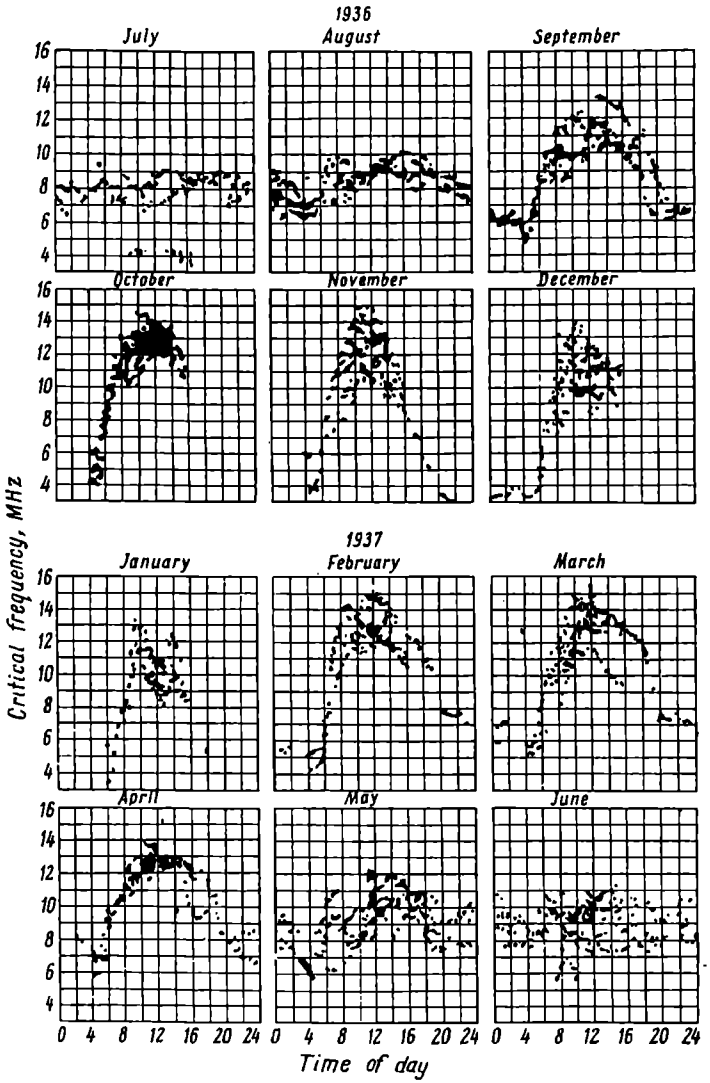


Fig. 4.12. Daily variations in critical frequencies for the *F2* layer between July, 1936 and June, 1937.

The diurnal variations in the electron density of the *F2* layer are subject to geomagnetic control, which is indicative of the contribution to the total ionization by corpuscular streams. There are also annual variations which manifest themselves in that the electron concentration increases in winter over the entire Northern Hemisphere.

In addition to the more stable regions and layers in the ionosphere, there is also what is known as the *Sporadic E* (*Es*) *layer* whose electron density may be up to ten times that of the normal *E* layer. The *Es* can occur at any hour of the day and in any season of the year. In temperate latitudes it occurs most often by day, in summer.

In the polar regions the *Es* occurs mainly at night and shows no pronounced seasonal variations. In the equatorial zone, the *Es* occurs predominantly by day. The probability of its occurrence increases with geomagnetic latitude. The *Es* is fairly stable with height, while its height differs from that of the normal *E* layer by not more than five to ten kilometres.

Numerous investigations have shown that the *Es* layer consists of relatively dense patches of electrons, that is, it has a lattice structure. The time during which the *Es* exists varies within broad limits, but does not usually exceed several hours. It is produced within a limited region of the ionosphere, and its horizontal extent is measured in hundreds or, probably, tens of kilometres. Sometimes, the *Es* is observed to drift in a horizontal direction at a speed of up to 300 kilometres per hour.

Existing hypotheses relate the formation of the *Es* to the infiltration of charged particles from the overlying *F2* layer due to the turbulent motion of air masses. A detailed description of the Sporadic *E* layer will be found in [71].

The outer region of the ionosphere is occupied by the radiation belts, discovered independently by Soviet and U.S. scientists through rocket launchings [72, 73]. These belts, known as the Van Allen belts, girdling the earth and consisting of the charged particles trapped by the terrestrial magnetic field have the shape of magnetic lines of force. The prevailing opinion is that there are three such belts arranged as shown in Fig. 4.13. The inner belt is situated in the interval of geomagnetic latitudes between $\pm(35°$ to $40°)$. In the plane of the geomagnetic equator, it begins at an altitude of about 600 kilometres in the Western Hemisphere and 1600 kilometres in the Eastern Hemisphere and extends out to a distance of about an earth radius. The inner zone is occupied by protons with an energy of 10^8 electron-volts and electrons with an energy of 10^6 electron-volts. These particles are presumably produced by the decay of the neutrons emitted by the earth's atmosphere under the action of cosmic rays.

The outer belt begins at an altitude of about 20,000 kilometres in the plane of the geomagnetic equator and extends out to a distance of seven earth radii. This belt has the shape of a horseshoe and comes within 300 to 1500 kilometres of the earth's surface in the interval of geomagnetic latitudes between $\pm$(55° to 70°). The outer belt appears to be filled with particles of solar origin energies from a few tens of kiloelectron-volts to megaelectron-volts.

The third radiation belt occurs in the interval of distances from 55,000 to 75,000 kilometres (8.5 to 11.5 earth radii). It is filled with

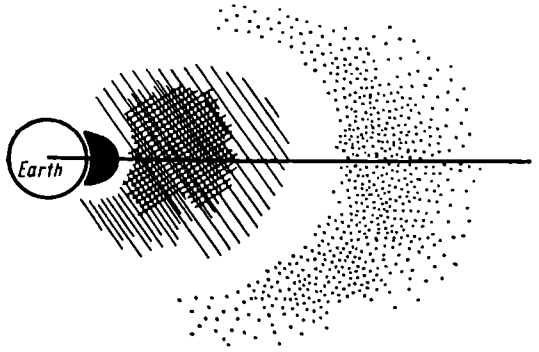


Fig. 4.13. Radiation belts around the earth

electrons of relatively low energies, but greater than 200 electron volts [74], [75], and [76]. Its size depends on the extent of geomagnetic disturbances.

Within the radiation belts, the charged particles are in oscillatory and rotational motions along and around the magnetic lines of force.

4.7. Radio Propagation in a Homogeneous Ionized Gas

In a homogeneous ionized gas of N electron number density in which an electron experiences ν collisions every second with neutral molecules, let there be a linearly polarized wave propagated in the direction of the X-axis, while the electric field vector **E** is oriented along the Z-axis (Fig. 4.14). In symbolic notation, the instantaneous electric field of the wave will be

$$E = E_m e^{i\omega t} \text{ V/m} \tag{4.40}$$

where ω is the radian frequency of the wave.

As a first approximation, let there be one electron of charge e and mass m in one cubic metre. The alternating electric field described

by (4.40) will act on the electron with a force given by

$$F = eE_m e^{i\omega t} \text{ N} \tag{4.41}$$

which will cause the electron to oscillate along the Z-axis.

At any instant, the force F acting on the electron will be balanced by the force of inertia, $m\frac{d^2z}{dt^2}$, and the frictional force due to the collision of the electron with neutral molecules. If we assume that in each collision the electron gives up to the molecule all the stored momentum $m\frac{dz}{dt}$, then the momentum will change by $\nu m\frac{dz}{dt}$ every second. On the other hand, the time rate of change of momentum is force—the force of friction in our case.

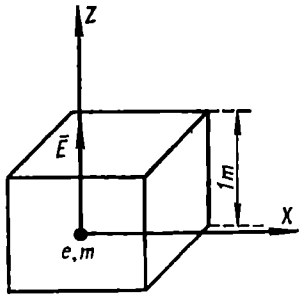


Fig. 4.14. Explaining radio propagation in a homogeneous ionized gas

Therefore, the motion of an electron may be described as

$$eE = m\frac{d^2z}{dt^2} + \nu m\frac{dz}{dt} \text{ N} \tag{4.42}$$

The steady-state solution of this equation is

$$z = z_m e^{i\omega t} \text{ m} \tag{4.43}$$

We seek to express the velocity of an electron, dz/dt, in terms of the time rate of change of the electric field, dE/dt. To do this, we write the acceleration in terms of velocity, and the electric field E in terms of its derivative. Then from Eqs. (4.40) and (4.43) we have

$$\frac{dE}{dt} = i\omega E \text{ V/ms} \tag{4.44}$$

and

$$\frac{d^2z}{dt^2} = i\omega\frac{dz}{dt} \text{ m/s}^2 \tag{4.45}$$

Substituting (4.44) and (4.45) in (4.42) gives

$$\frac{e}{i\omega}\frac{dE}{dt} = m\frac{dz}{dt}(\nu + i\omega) \text{ N} \tag{4.46}$$

Solving this algebraic equation for dz/dt gives

$$\frac{dz}{dt} = \frac{e}{i\omega m(\nu + i\omega)}\frac{dE}{dt} = \left[-\frac{e}{m(\omega^2 + \nu^2)} - i\frac{\nu e}{\omega m(\omega^2 + \nu^2)}\right]\frac{dE}{dt} \text{ m/s} \tag{4.47}$$

Actually, every cubic metre contains N electrons each of which has a velocity given by (4.47). The total charge of unit volume is Ne. By definition, the instantaneous current density is the quantity

of electricity passing through unit cross-sectional area in unit time, assuming that in the meantime the electrons are moving at a constant speed equal to the instantaneous value. Under these conditions, there will be dz/dt volume units passing in one second in the direction of the Z-axis; that is, the convection-current density will be

$$j_c = Ne\frac{dz}{dt} = \left[-\frac{Ne^2}{m(\omega^2+\nu^2)} - i\frac{Ne^2\nu}{\omega m(\omega^2+\nu^2)} \right]\frac{dE}{dt} \text{ A/m}^2 \quad (4.48)$$

In the air (irrespective of its ionization), an alternating electric field gives rise to a displacement current whose density is given by

$$j_d = \varepsilon_0\frac{dE}{dt} \text{ A/m}^2 \quad (4.49)$$

The total current density will be

$$j = j_c + j_d = \left[\varepsilon_0 - \frac{Ne^2}{m(\omega^2+\nu^2)} - i\frac{Ne^2\nu}{m\omega(\omega^2+\nu^2)} \right]\frac{dE}{dt} \text{ A/m}^2 \quad (4.50)$$

The term in square brackets resembles the complex permittivity of an imperfectly conducting medium (see § 1.3). From a comparison of this term with the expression for the complex permittivity

$$\varepsilon_k = \varepsilon - i\frac{\sigma}{\omega} \text{ F/m} \quad (4.51)$$

we may conclude that an ionized gas behaves like an imperfectly conducting medium with the characteristics

$$\varepsilon_i = \varepsilon_0 - \frac{Ne^2}{m(\omega^2+\nu^2)} \text{ F/m} \quad (4.52)$$

and

$$\sigma_i = \frac{Ne^2\nu}{m(\omega^2+\nu^2)} \text{ Siemens/m} \quad (4.53)$$

Substituting the numerical values of the mass and charge of an electron and also the permittivity of free space gives

$$\varepsilon_i' = \varepsilon_i/\varepsilon_0 = 1 - \frac{Ne^2}{\varepsilon_0 m(\omega^2+\nu^2)} = 1 - 3190\frac{N}{\omega^2+\nu^2} \quad (4.54)$$

$$\sigma_i = 2.82 \times 10^{-8}\frac{N\nu}{\omega^2+\nu^2} \text{ Siemens/m} \quad (4.55)$$

where the subscript i indicates that the characteristics apply to an ionized gas, and ε_i', as always, stands for the relative (dimensionless) permittivity.

In cases where $\omega^2 \gg \nu^2$, the expressions reduce to

$$\varepsilon_i' \cong 1 - 80.8N/f^2 = 1 - 80.8N_{(\text{cm}^{-3})}/f^2\,(\text{kHz}) \quad (4.54a)$$

$$\sigma_i \cong 2.82 \times 10^{-8}N\nu/\omega^2 \text{ Siemens/m} \quad (4.55a)$$

In the limiting case of $\nu = 0$, that is, when no collisions occur, the permittivity is expressed by (4.54a), and the conductivity reduces to zero. As should be expected, in the absence of collisions an ionized gas behaves like an ideal dielectric.

Conversely, if $\omega^2 \ll \nu^2$, then

$$\varepsilon_i' \cong 1 - 3190N/\nu^2 \tag{4.54b}$$

and

$$\sigma_i \cong 2.82 \times 10^{-8} N/\nu \text{ Siemens/m} \tag{4.55b}$$

So far we have only dealt with electrons and ignored the fact that each unit volume of air contains a number of ions, too. Consider the case of simple recombination, when there are equal numbers of electrons and ions present.

It is obvious that all that has been derived for the density of electron flow fully applies to that of ion flow. It is only necessary to use the mass of an ion, m_i, instead of that of an electron, m, in Eq. (4.54). Since positive ions move in an opposite direction to electrons, and their charges are unlike, their flows must be added together. Then (4.54) takes the form

$$\varepsilon_i' = 1 - \frac{Ne^2}{\varepsilon_0 m\omega^2} - \frac{Ne^2}{\varepsilon_0 m_i\omega^2} \tag{4.56}$$

Since the mass of an atom of nitrogen, one of the most easily ionizable gases, is 25,800 times that of an electron, we may well ignore the effect of ions.

In the more general case of multi-step recombination, when

$$N + N_- = N_+ \text{ m}^{-3} \tag{4.57}$$

where N_- and N_+ are the numbers of negative and positive ions per unit volume, respectively, Eq. (4.56) takes the form

$$\varepsilon_i' = 1 - \frac{Ne^2}{\varepsilon_0 m\omega^2} - \frac{N_-e^2}{\varepsilon_0 m_i\omega^2} - \frac{N_+e^2}{\varepsilon_0 m_i\omega^2} \tag{4.56a}$$

In cases where $N_- \gg N$, the effect of ions cannot be neglected, and it is important to assess the respective contributions of electrons and ions in each particular case.

4.8. The Phase and Group Velocities of Radio Waves in an Ionized Gas

With the collisions between electrons and neutral molecules neglected, we may, in a first approximation, treat an ionized gas as perfect dielectric the relative permittivity of which is given by

$$\varepsilon_i' = 1 - 80.8N/f^2 \tag{4.54a}$$

This expression is unusual in more than one respect. For one thing, the absolute value of the permittivity appears to be less than unity which in itself is unusual, since all solids, liquids and gases have a permittivity of more than unity, and it is only for a vacuum that ε' is deemed to be unity. For another, the permittivity is seen to be frequency-dependent. At a certain value of f it reduces to zero, while at the lower frequencies it becomes negative. It is easy to realize that these anomalies stem from the fact that we have added together the convection current due to the electric field of the wave and the displacement current and considered the total current as an effective displacement current of a dielectric medium.

From the field theory it is known that in a pure dielectric radio waves are propagated at the velocity given by

$$u_p = c/\sqrt{\varepsilon_i'} \text{ m/s} \tag{4.58}$$

which is defined as the velocity of travel of an equiphase surface in the direction of the wave normal and is called the phase velocity of the wave. In our case, $\varepsilon_i' < 1$, and so the phase velocity exceeds that of light in free space. Besides, the phase velocity varies with frequency. The media in which the velocity of waves varies with frequency are called *dispersive* because this results in dispersion of the composite signal, that is, the separation of its component frequencies.

According to the Fourier integral theorem, such a signal can be represented by an infinite sum of harmonic components occupying a continuous frequency spectrum from zero to infinity. In a non-dispersive medium, all harmonic components, irrespective of their individual frequencies, are propagated with the same velocity, and cover the same path and, on being combined at the point of reception, will reproduce the original waveform of the pulse. That is, the group of waves is propagated without distortion under such conditions.

In a dispersive medium the harmonic components travel with different velocities and, on being combined at the point of reception, will produce a pulse whose waveform will be different from the original one. Thus, the signal is distorted in transmission.

Therefore, the velocity of a composite wave has to be defined in one more way so as to take care of this distortion. This is done by introducing the concept of group velocity. It is defined as the velocity of propagation of the crest of a group of interfering waves where the component wave trains have slightly different individual frequencies and phase velocities.

To determine the group velocity of radio waves in an ionized gas, it will suffice to consider a simple group composed of two harmonic

components very near in frequency. Assuming that the amplitudes of the elementary waves are equal, we have

$$E_1 = E_m \cos \omega_1 (t - x/u_1) \text{ V/m}$$

$$E_2 = E_m \cos \omega_2 (t - x/u_2) \text{ V/m}$$

The resultant field is then given by

$$E = 2E_m \cos \frac{1}{2} \left[(\omega_1 - \omega_2) t - \left(\frac{\omega_1}{u_1} - \frac{\omega_2}{u_2} \right) x \right] \times$$

$$\times \cos \frac{1}{2} \left[(\omega_1 + \omega_2) t - \left(\frac{\omega_1}{u_1} + \frac{\omega_2}{u_2} \right) x \right] \text{ V/m} \qquad (4.59)$$

At a fixed instant, the addition of the two frequencies expressed by (4.59) results in beats, as shown in Fig. 4.15. The dotted line

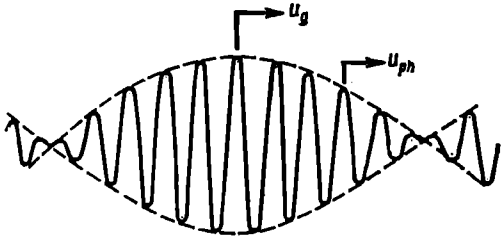


Fig. 4.15. Explaining the group velocity of radio waves

is the envelope of the beats. It is represented by the first term in (4.59), and the r.f. modulated wave, by the second.

The energy of the wave is a maximum at the crest of the envelope, because the amplitude of the wave is highest exactly there. Consequently, in order to find the group velocity, it will suffice to determine the velocity of travel for the crest of the envelope on the assumption that $\omega_1 \to \omega_2$, for this velocity must be determined for a certain definite frequency.

The condition for the envelope to be a maximum is described by

$$(\omega_1 - \omega_2) t - (\omega_1/u_1 - \omega_2/u_2) x = \pi n \qquad (4.60)$$

where n is an integer.

Differentiating and passing to the limit, we get

$$u_g = \lim_{\omega_1 \to \omega_2} (dx/dt) = \lim_{\omega_1 \to \omega_2} \frac{\omega_1 - \omega_2}{\omega_1/u_1 - \omega_2/u_2} = \frac{d\omega}{d(\omega/u)} \qquad (4.61)$$

For a perfect dielectric, (4.58) may be re-written as

$$u = c/n \text{ m/s} \qquad (4.58a)$$

where $n = \sqrt{\varepsilon_i'}$ is the refractive index.

Substituting (4.58a) in (4.61) gives

$$u_g = \frac{c\,d\omega}{\omega\,dn + n\,d\omega} = \frac{c}{n + f\dfrac{dn}{df}} \text{ m/s} \tag{4.62}$$

The condition for the r.f. modulated wave to be a maximum is given by

$$(\omega_1 + \omega_2)\,t - \left(\frac{\omega_1}{u_1} + \frac{\omega_2}{u_2}\right) x = \pi n \tag{4.63}$$

Differentiating, passing to the limit, and noting that the velocity of travel of this maximum is the phase velocity of the wave, we get

$$u_p = \lim_{\omega_1 \to \omega_2} dx/dt = \lim_{\omega_1 \to \omega_2} \frac{\omega_1 + \omega_2}{\omega_1/u_1 + \omega_2/u_2} = u = c/n \text{ m/s} \tag{4.64}$$

Equation (4.64) confirms that the quantity u in the original expression does represent the phase velocity of the wave. On the other hand, Eq. (4.62) clearly shows that the group velocity u_g differs from the phase velocity u_p only when n is frequency-dependent and when the derivative dn/df is finite.

In an ionized gas and in the absence of collisions, we may write on the basis of (4.54a) that

$$n = \sqrt{\varepsilon_i'} = \sqrt{1 - 80.8N/f^2} \tag{4.54b}$$

Basing ourselves on this expression, we can readily prove that the product of the phase velocity and the group velocity of a wave is equal to the square of the velocity of light in a vacuum.

Multiplying (4.62) by (4.64) gives

$$u_p \cdot u_g = \frac{c^2}{n^2 + \dfrac{f}{2}\dfrac{d(n^2)}{df}} \text{ m}^2/\text{s}^2 \tag{4.65}$$

Substituting the expression for n^2 from (4.54b), we get

$$u_p \cdot u_g = \frac{c^2}{1 - 80.8\dfrac{N}{f^2} + \dfrac{f}{2}80.8 \times 2N/f^3} = c^2 \text{ m}^2/\text{s}^2 \tag{4.65a}$$

Among other things, from Eq. (4.65a) it follows that

$$u_g = cn \text{ m/s} \tag{4.66}$$

Since in a homogeneous ionized gas $n < 1$ always, the group velocity is lower than the velocity of light in free space also always.

As regards ionospheric propagation, knowledge of the phase velocity is important in dealing with refraction and reflection, for the wave trajectory is in the long run decided by the phase velocity. Knowledge of the group velocity is important in measuring the

delay of radio waves reflected from the ionosphere. Practical knowledge of the group velocity comes in useful in processing the measurements taken by ionospheric stations.

Numerical calculations show that for signals with a duration of a few milliseconds the distortion caused by the dispersive properties of the ionosphere may be neglected.

4.9. Radio Propagation in a Homogeneous Ionized Gas in the Presence of a Permanent Magnetic Field

In dealing with ionospheric propagation, it is to be remembered that radio waves are propagated in an ionized gas within the magnetic field of the earth. The average strength of the terrestrial magnetic field is about 40 amperes per metre, which works out to a magnetic induction of 0.5 gauss in the emu units.

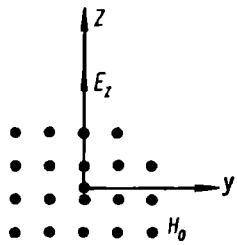


Fig. 4.16. Explaining the concept of gyrofrequency

The permanent magnetic field substantially affects radio propagation and makes the ionized gas an anisotropic medium, that is,one having different properties in differing directions.

Above all, the earth's constant magnetic field causes the electrons to follow complicated trajectories instead of oscillating rectilinearly. Referring to Fig. 4.16, there is a system of coordinates. As before, radio waves are propagated along the X-axis, and the electric field is oriented along the Z-axis, indicated by the component E_z. What is new is that there is a constant magnetic field, H_0, the lines of force of which are parallel with the X-axis (in the sketch they are shown by dots).

As soon as the electric field brings up an electron to a velocity **u** (originally, along the Z-axis), the magnetic field exerts the Lorentz force on it

$$\mathbf{F}_m = e\mu_0 [\mathbf{u}\mathbf{H}_0] \text{ N} \tag{4.67}$$

From (4.67) it follows that the Lorentz force is normal to the velocity vector **u** and the magnetic field vector $\mathbf{H}_0$. In our case, the velocity vector **u** is normal to the vector $\mathbf{H}_0$, and so the absolute value of F_m is given by

$$F_m = e\mu_0 u H_0 \text{ N} \tag{4.67a}$$

If we assume that after an electron has been brought up to velocity u the wave field vanishes, then the transverse force F_m will cause the electron to trace out a curved path whose radius can be found from the equality of the centrifugal force and the Lorentz force.

On neglecting the collisions between electrons and neutral molecules, the equation of the motion of an electron takes the form

$$mu^2/R = e\mu_0 u H_0 \text{ N} \tag{4.68}$$

Equation (4.68) shows above all that, neglecting the field due to the radio wave, the electrons will move in a circle of the radius

$$R = mu/e\mu_0 H \text{ m} \tag{4.68a}$$

which is directly proportional to the initial velocity of the motion. It is very important that the period of one orbit

$$T = 2\pi R/u = 2\pi m/e\mu_0 H_0 \text{ s} \tag{4.69}$$

is independent of the initial velocity and is a constant for a given strength of the earth's magnetic field, or, in terms of frequency

$$1/T = f_m = e\mu_0 H_0/2\pi m \text{ Hz} \tag{4.70}$$

which is likewise a constant for a given strength of the earth's magnetic field. Thus, the rotation of electrons under the effect of the terrestrial magnetic field exhibits resonance characteristic. This is why it is sometimes referred to as gyromagnetic resonance, and the respective frequency is called the gyro-frequency.

Substituting the numerical values of e and m (in coulombs and kilograms, respectively), μ_0, and the average value of $H_0 = 40$ amperes per metre in (4.70) gives

$$f_m \cong 1.4 \text{ MHz}$$

If the radio wave continues after the electron has been set in motion, its path will naturally acquire a more complicated form. It is important to note that this path retains elements of rotational motion in any circumstances. Because of this, the propagation of a plane wave may produce additional (transverse and longitudinal) components of the electric field which, given certain conditions, may lead to the rotation of the plane of polarization and double refraction, that is, processes associated with anisotropic media.

On the other hand, if the frequency of the wave is the same as the gyro-frequency, the motion of the electron, supported by the wave field, will be along an unfolding spiral. Owing to collisions with neutral molecules, this motion will be interrupted and resumed again repeatedly. Under such conditions the electrons will have a greater mean velocity than they do in the absence of a magnetic field, and this will lead to increased losses in collisions. It may be expected that waves at a frequency of about 1.4 megahertz will experience increased attenuation in the ionosphere.

To formulate a general theory describing radio propagation in an ionized gas in the presence of a constant magnetic field, let us introduce a Cartesian system of coordinates and assume that a plane radio wave is propagated along the X-axis (Fig. 4.17). Without any loss of generality, we may take it that the constant magnetic field vector H_0 is in the XOZ plane, making an angle β with the OX axis. Its longitudinal component is $H_l = H_{OX} = H_0 \sin \beta$, and its transverse component is $H_t = H_{OZ} = H_0 \cos \beta$. The component along the OY axis is zero, $H_{OY} = 0$.

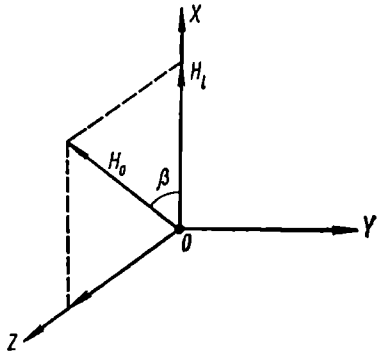


Fig. 4.17. The coordinate system used in studies into effects of the terrestrial field on radio propagation

Neglecting the collisions between electrons and neutral molecules and, as a consequence, losses in propagation, and assuming that the field is a harmonic function of time, we seek to determine the refractive index n of the ionized gas for the radio wave propagated along the OX axis, whose transverse components are described as

$$E_y = E_{my} e^{-i\frac{\omega}{c} nx} \text{ V/m}$$
$$E_z = E_{mz} e^{-i\frac{\omega}{c} ny} \text{ V/m} \qquad (4.71)$$

Analytically it has been shown [68] that the refractive index is

$$n^2_{1,2} = 1 - \frac{\omega_0^2}{\omega^2 - \frac{\omega^2\omega_l^2}{2(\omega^2-\omega_0^2)} \mp \sqrt{\left[\frac{\omega^2\omega_l^2}{2(\omega^2-\omega_0^2)}\right]^2 + \omega_l^2\omega^2}} \qquad (4.72)$$

The amplitude ratio of the transverse components of the field is

$$\left(\frac{E_{my}}{E_{mz}}\right)_{1,2} = i\left\{\frac{\omega_l^2\omega}{2(\omega^2-\omega_0^2)\omega_l} \mp \sqrt{\left[\frac{\omega_l^2\omega}{2\omega_l(\omega^2-\omega_0^2)}\right]^2 + 1}\right\} \qquad (4.73)$$

The longitudinal field component is described by

$$E_x = \frac{\omega\omega_t\omega_0^2}{(\omega^2-\omega_0^2)(\omega_l^2-\omega^2)+\omega^2\omega_t^2}\left(-iE_y + \frac{\omega_l}{\omega}E_z\right) \text{ V/m} \qquad (4.74)$$

where

$$\left.\begin{aligned} \omega_0^2 &= Ne^2/\varepsilon_0 m = \text{plasma frequency of the ionised gas, Hz} \\ \omega_t &= \omega_m \sin\beta = \mu_0 e H_t/m = \text{transverse gyro-frequency, Hz} \\ \omega_l &= \omega_m \cos\beta = \mu_0 e H_l/m = \text{longitudinal gyro-frequency, Hz} \end{aligned}\right\} \qquad (4.75)$$

Among other things, from Eqs. (4.75) it follows that

$$\omega_m = \mu_0 e H_0/\text{m Hz}$$

because

$$H_0 = \sqrt{H_l^2 + H_t^2}\ \text{A/m}$$

The equations derived above show that in the presence of a constant magnetic field H_0, an ionized gas behaves like an anisotropic medium for radio waves. Formally, this manifests itself in that the permittivity acquires the properties of a tensor. The fact that there are two signs in (4.72) for the refractive index and in (4.73) for the kind of polarization indicates that the radio wave experiences double refraction. The subscript *1* applies to the extraordinary ray, and the subscript *2* to the ordinary ray (to be defined a bit later).

Consider some special cases.

There is no constant magnetic field. Under these conditions, $H_l = H_t = 0$, whence it immediately follows that $\omega_l = \omega_t = 0$, and Eq. (4.54) reduces to the usual expression for the refractive index of an ionized gas

$$n^2 = \varepsilon_l' = 1 - \omega_0^2/\omega^2 = 1 - Ne^2/\varepsilon_0 m\omega^2 \tag{4.76}$$

The ionized gas behaves like an isotropic medium.

The radio wave is propagated in a longitudinal magnetic field. In this case, $H_t = 0$ and $\omega_t = 0$. Substituting these values in (4.72) gives

$$n_{1,2}^2 = 1 - \frac{\omega_0^2}{\omega(\omega \mp \omega_l)} \tag{4.77}$$

On the other hand, substituting $\omega_t = 0$ in (4.73) gives

$$(E_y/E_x)_{1,2} = \pm i \tag{4.78}$$

Equation (4.78) describes circularly polarized radio waves. The upper sign applies to the resultant vector rotating clockwise (when viewed in the direction of propagation), and the lower sign to the resultant vector rotating counter-clockwise. As a consequence, the refractive index given by (4.77) likewise applies to clockwise and counter-clockwise circularly polarized waves.

When $\omega_t = 0$, (4.74) gives $E_x = 0$, which reveals the absence of a longitudinal component in the electric field.

If at some point in space the electric field of a plane polarized field is given by

$$E_z = E_m \cos \omega t\ \text{V/m}$$

this field may be represented as being due to two circularly polarized waves

$$\left.\begin{aligned} E_z &= \frac{E_m}{2}\cos\omega t \quad \text{V/m} \\ E_y &= \frac{E_m}{2}\sin\omega t \quad \text{V/m} \end{aligned}\right\} \text{(A)}$$
$$\left.\begin{aligned} E_z &= \frac{E_m}{2}\cos\omega t \quad \text{V/m} \\ E_y &= -\frac{E_m}{2}\sin\omega t \quad \text{V/m} \end{aligned}\right\} \text{(B)} \qquad (4.79)$$

In the absence of an external magnetic field, both waves, A and B, are propagated with the same velocity. At any point in space they will therefore combine to produce a plane polarized wave with the plane of polarization oriented as before.

In our case, wave A is propagated with a velocity $u_2 = c/n_2$, and wave B, with a velocity $u_1 = c/n_1$. While at $x = 0$ the resultant field is oriented in the direction of the Z-axis, at distance x the components along the Z- and X-axes will be given by

$$\left.\begin{aligned} E_z &= E_m \cos\left[\frac{\omega x}{2c}(n_2 - n_1)\right]\cos\omega\left[t - \frac{x}{2c}(n_2 + n_1)\right] \\ E_y &= E_m \sin\left[\frac{\omega x}{2c}(n_2 - n_1)\right]\cos\omega\left[t - \frac{x}{2c}(n_2 + n_1)\right] \end{aligned}\right\} \qquad (4.80)$$

Equations (4.80) show that the two components are in phase at any point in space. In other words, the wave remains linearly polarized. The plane of polarization, however, will be oriented differently. The angle through which the plane of polarization will have to be rotated on passing the distance x can be found from the expression

$$\tan\theta = E_y/E_z = \tan\frac{\omega x}{2c}(n_2 - n_1)$$

whence

$$\theta = \frac{\omega x}{2c}(n_2 - n_1) \qquad (4.81)$$

Denoting the distance after which the plane of polarization is rotated through 360° as $x_{2\pi}$ and substituting $\theta = 2\pi$ in (4.81) gives

$$x_{2\pi} \cong \frac{2\lambda_0}{n_2 - n_1} \text{ m} \qquad (4.82)$$

where λ_0 is the wavelength in a vacuum.

For short waves, the expressions for n_1 and n_2 may be considerably simplified, and Eq. (4.82) may be re-written as

$$x_{2\pi} \cong \frac{2\lambda_0\omega^3}{\omega_0^2\omega_l} \text{ m} \qquad (4.82a)$$

It is not difficult to realize the physical implications of the rotation of the plane of polarization. The point is that electrons propagated in a longitudinal magnetic field are set into a rotational motion in a plane normal to the direction of propagation. As the magnetic field splits the linearly polarized wave into two components (called an ordinary wave and an extraordinary wave) rotating in opposite directions, the direction of rotation of the resultant field will be in the direction of rotation of the electrons only for one of these two components. Naturally, this component will be propagated differently (above all, in terms of velocity) than the component for which the direction of rotation of the field vector is opposite to that of rotation of the electrons. Obviously, the difference in velocity of propagation must result in the rotation of the plane of polarization.

The wave is propagated in a transverse magnetic field. Now, $H_l = 0$, and $\omega_l = 0$. Substituting these values in (4.72) gives

$$n_1^2 = 1 - \frac{\omega_0^2}{\omega^2 - \dfrac{\omega^2\omega_l^2}{\omega^2 - \omega_0^2}} \tag{4.83}$$

$$n_2^2 = 1 - \frac{\omega_0^2}{\omega^2} \tag{4.84}$$

It is to be noted that the refractive index n_2 has the same value as in the absence of an external magnetic field.

To determine the kind of polarization of the wave, let us refer to Eq. (4.73). Substituting $\omega_l = 0$ gives

$$\left| \left(\frac{E_{my}}{E_{mz}} \right)_1 \right| = \infty$$

which corresponds to a case where $E_{my} \neq 0$ and $E_{mz} = 0$, and

$$\left| \left(\frac{E_{my}}{E_{mz}} \right)_2 \right| = 0,$$

which corresponds to a case where $E_{my} = 0$ and $E_{mz} \neq 0$.

Thus, the refractive index n_1 describes the propagation of a linearly polarized wave which is propagated at right angles to the direction of the magnetic field and whose electric field vector E_{my} is likewise at right angles to the magnetic field (Fig. 4.18). Similarly, the refractive index n_2 describes the propagation of a wave which is propagated at right angles to the magnetic field but whose electric field vector is parallel to the lines of force of the constant magnetic field (Fig. 4.18).

As is customary in optics, the linearly polarized component E_{mz} which is not affected by an external magnetic field is called an *ordinary ray*, and the linearly polarized component E_{my} affected

by the external magnetic field is called an *extraordinary ray.*

Substituting $\omega_t = 0$ in (4.74), the longitudinal component of the electric field is given by

$$E_x = -i \frac{\omega \omega_t \omega_0^2}{\omega^2 \omega_t^2 - \omega^2 (\omega^2 - \omega_0^2)} E_y \tag{4.85}$$

which shows that the longitudinal component is an attribute of the extraordinary ray only.

From the foregoing we may conclude that a linearly polarized ray whose electric field component E_m makes an angle α with the direction of the magnetic field is split into two linearly polarized rays, ordinary and extraordinary. The ordinary ray is propagated as if there were no magnetic field present. Conversely a constant magnetic field strongly affects the propagation of the extraordinary ray by changing the velocity of propagation given by $u_1 = c/n_1$ where n_1 is in turn given by (4.83). Besides, the extraordinary ray has a longitudinal electric-field component in quadrature with the transverse field component.

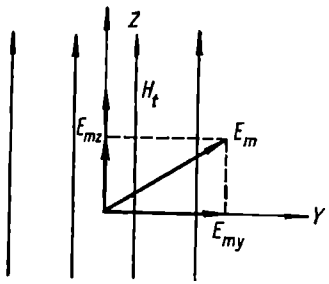


Fig. 4.18. Radio propagation in a transverse magnetic field

If the original electric field is oriented in the direction of the Z- or Y-axis, the incident wave will not be split into two—the wave will be propagated as either an ordinary ray or an extraordinary ray.

It is not difficult to understand physically what happens as a wave is propagated in a transverse magnetic field. If the electric field is oriented along the Z-axis, the electron gains speed in the same direction and, according to (4.67), the Lorentz force F_m reduces to zero. That is, the magnetic field has no effect on the propagation of such waves.

If the electric field is oriented along the Y-axis, the electrons are set into a rotational motion in the XY plane which, firstly, affects the velocity of wave propagation and, secondly, produces a longitudinal component in the electric field.

Wave propagation in the general case. In the most general case both ω_t and ω_l are other than zero, and we have to use the complete equations.

Equation (4.73) shows that both the ordinary and the extraordinary rays are then elliptically polarized. Setting

$$\frac{\omega_t^2 \omega}{2\omega_l (\omega^2 - \omega_0^2)} = s \tag{4.86}$$

we may re-write (4.73) as

$$\left(\frac{E_{my}}{E_{mz}}\right)_{1,\,2} = i\,(s \pm \sqrt{s^2+1}) \tag{4.87}$$

Hence,

$$\left(\frac{E_{my}}{E_{mz}}\right)_1 \left(\frac{E_{my}}{E_{mz}}\right)_2 = 1 \tag{4.88}$$

or, which is the same thing,

$$\left(\frac{E_{my}}{E_{mz}}\right)_1 = \left(\frac{E_{mz}}{E_{my}}\right)_2 = i\beta \tag{4.89}$$

where $\beta = s + \sqrt{s^2+1}$.

Since the phase shift between E_y and E_z is 90°, we may say that the axes of the ellipses of polarization are coincident with the Y- and Z-axes. On the other hand, (4.89) clearly indicates that the major axes of the ellipses of polarization are mutually perpendicular. Lastly, the fact that for the extraordinary ray the right-hand side of (4.87) is positive (the upper sign) while for the ordinary ray it is negative (the lower sign) indicates that the directions of rotation for the resultant vectors are opposite in the two cases.

The overall conclusion is that in the general case a linearly polarized ray is split into two elliptically polarized rays.

Owing to (4.74), each of the two elliptically polarized rays has a longitudinal electric-field component. These components may be found by (4.74) and (4.89). Setting

$$\frac{\omega_0^2}{(\omega^2-\omega_0^2)(\omega_l^2-\omega^2)+\omega^2\omega_l^2} = q\,s^2 \tag{4.90}$$

we get the following:

for the extraordinary ray

$$E_{mx1} = q\omega_l\,(\omega s + \omega_l)\,E_{mz1} \tag{4.91a}$$

for the ordinary ray

$$E_{mx2} = q\omega_l\,\frac{\omega - s\omega_l}{s}\,E_{mz2} \tag{4.91b}$$

Equations (4.90) and (4.91) show that the longitudinal electric-field components of both the extraordinary and ordinary rays are in phase with the transverse component oriented in the direction of the magnetic field.

Thus, we may conclude that when a linearly polarized wave is propagated in an arbitrary direction relative to the direction of a constant magnetic field, the wave is split into two elliptically polarized rays with longitudinal electric-field components described by the triplets (E_{1y}, E_{1z}, E_{1x}) and (E_{2y}, E_{2z}, E_{2x}), respectively.

These rays are propagated with different velocities ($u_1 = c/n_1$ and $u_2 = c/n_2$), the major axes of the ellipses of polarization are at right angles to each other, and the resultant vectors rotate in opposite directions.

For simplicity, we have so far ignored the collisions between electrons, neutral molecules and ions. In taking them into account, the already complicated expressions for the refractive index grow still more unwieldy. Besides, in dealing with collisions, we must take into account the related losses, that is, the attenuation of radio waves.

As with radio propagation in an imperfectly conducting medium, the attenuation experienced by radio waves in an ionized gas in the presence of a constant magnetic field due to the absorption of wave energy is described by the absorption coefficients δ_1 and δ_2, the former applying to the extraordinary and the latter, to the ordinary ray.

4.10. Refraction and Reflection of Radio Waves in the Ionosphere

Reflection from an Ionized Layer. So far we have dealt with radio propagation in a homogeneous ionized gas. The real ionosphere is an inhomogeneous ionized gas. The inhomogeneity of the ionosphere manifests itself above all in that radio waves follow curved and not straight paths (as in all the cases discussed so far). Given certain conditions, radio waves can be fully reflected from the ionosphere and come back to the earth.

To begin with, we shall consider radio propagation in a "plane" ionosphere, such that the surfaces of equal electron number densities are planes parallel with one another. Furthermore, this ionosphere is assumed to consist of thin plane strata within each of which the electron number density is constant. Denoting the electron number density in the first stratum as N_1, in the second as N_2, and so on, we deem that $0 < N_1 < N_2 < \ldots < N_n < N_{n+1}$. Furthermore, we shall assume that the ray of frequency f emerging from the unionized atmosphere is incident upon the lowest ionized stratum at an angle φ_0 (Fig. 4.19).

Using the expression

$$n = \sqrt{1 - 80.8N/f^2}$$

for the refractive index, we may write

$$1 > n_1 > n_2 > \ldots > n_n > n_{n+1}$$

Applying Snell's law to each interface gives

$$1 \sin \varphi_0 = n_1 \sin \varphi_1 = \ldots = n_n \sin \varphi_n \qquad (4.102)$$

After a sufficient number of reflections, the angle of incidence at the nth stratum may approach 90° very nearly. Putting $\varphi_n \cong 90°$ in (4.92) and retaining the extreme terms in the equality, we get

$$\sin \varphi_0 = n_n \tag{4.93}$$

Equation (4.93) shows that at the nth stratum the ray is brought to the horizontal above the earth.

It remains to determine why a radio wave is turned back at the crest of its path.

The physical cause of this is total internal reflection. As will be recalled, total reflection occurs when electromagnetic energy passes

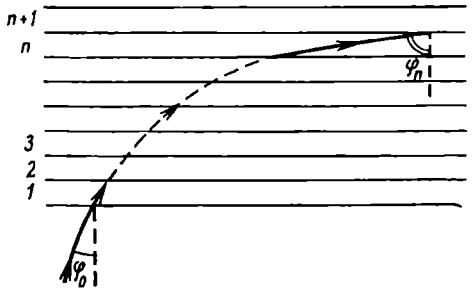


Fig. 4.19. Refraction of radio waves in a plane stratified ionosphere

from a medium with a greater density to one which has a lower density at a certain critical angle of incidence such that

$$\varphi_c = \arcsin \frac{n_{n+1}}{n_n} \tag{4.94}$$

Assuming that the refractive indices of the nth and $(n+1)$th strata are related as

$$n_{n+1} = n_n - \Delta n$$

the condition for the ray path to be turned back towards the earth is given by

$$\varphi_n > \arcsin (1 - \Delta n/n_n) \tag{4.95}$$

Actually, the refractive index of the ionosphere varies with height gradually, and so (4.95), which holds solely for a plane stratified ionosphere, must be replaced with another equation.

To derive the actual condition for the ray path to be turned back towards the earth, we may depart from the intuitive (albeit obvious) assumption that a radio wave can come back to the earth only if its radius of curvature at the crest of the path is less than $(a + h)$ where a

is the earth's radius, and h is the height of the turning point above the earth. Otherwise, as can be seen from Fig. 4.20, the radio wave even reflected from the ionosphere (the path MN) will not come back to the earth.

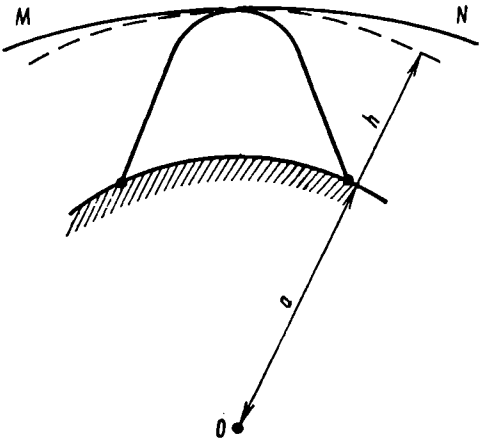


Fig. 4.20. Return of a radio wave to earth

As is shown in § 3.3, the radius of curvature of the ray path in a stratified atmosphere is given by (3.10) which we write again for convenience

$$R = \frac{n}{\sin\varphi\,(-dn/dh)}\ \text{m} \qquad (4.96)$$

where n is the refractive index, and φ is the angle between the path element and the vertical. Of course, (4.96) applies equally to the troposphere and the ionosphere.

At the crest of the path, $\varphi = 90°$, and (4.96) reduces to

$$R = n/(-dn/dh)\ \text{m} \qquad (4.96a)$$

The refractive index of the ionosphere is given by

$$n = \sqrt{\varepsilon_i'} = \sqrt{1 - 80.8N/f^2} \qquad (4.97)$$

Substituting (4.96a) in the condition for the ray path to be turned back to earth

$$R < (a+h) \qquad (4.98)$$

we may re-write (4.98) as

$$dN/dh > \frac{2n^2f^2}{80.8\,(a+h)}\ \text{m}^{-4} \qquad (4.99)$$

Equation (4.99) shows that at the turning point the electron number density should increase with height and that the vertical gradient of the electron number density should be not less than some critical value. The absolute value of this gradient is found in the example that follows.

Example 4.2. Determine the condition for the ray path to turn back at $f = 30$ MHz on the assumption that the highest point of the path is at $h = 200$ km above the earth's surface.

Solution: The statement of the problem describes the most unfavourable case because the frequency enters the numerator on the right-hand side of Eq. (4.99).

Substituting the numerical value of the earth's radius, $a = 6.37 \times 10^6$ m, in the equation, we get

$$dN/dh > 27 \times 10^5 \text{ m}^{-4}$$

However, it is more convenient to use the electron number density per cubic metre. Then

$$dN/dh > 2.7 \text{ cm}^{-3}\text{m}^{-1}$$

that is, for the ray path to turn back it is sufficient for the electron number density to increase by as little as 3 electrons per cubic centimetre for a change of one metre in height at the highest point of the ray path.

Thus, the additional condition for the ray to return to earth is satisfied automatically nearly always, and the decisive condition is (4.93).

The additional condition implies the following: (1) at the point of reflection the electron number density must increase; (2) reflection cannot occur in a region where the electron number density is a maximum, and (3) nor can this happen in a region where the electron number density decreases with height.

Sometimes, it may be convenient to give this condition a slightly different form.

Substituting the expression for n of an ionized gas in (4.93), we may write

$$\sin \varphi_0 = \sqrt{1 - 80 \cdot 8N/f^2} \tag{4.93a}$$

Setting $\varphi_0 = 0$ in (4.93a), we get the condition for a vertical ray of frequency f_{vert} to return to earth from the same stratum of electron density N, namely

$$0 = \sqrt{1 - 80.8N/f_{vert}^2} \tag{4.93b}$$

From (4.93a) and (4.93b) we get

$$f = f_{vert} \sec \varphi_0 \text{ Hz} \tag{4.100}$$

which relates an oblique ray of frequency f and a vertical ray of frequency f_{vert} that can be returned to earth from one and the same layer in the ionosphere. Equation (4.100) is sometimes called the *secant law*.

Maximum Usable and Critical Frequencies. Equation (4.93) shows that if the maximum electron density in a given region is

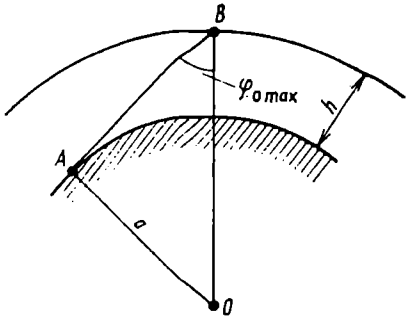


Fig. 4.21. Maximum angle of incidence at the entrance to the ionosphere

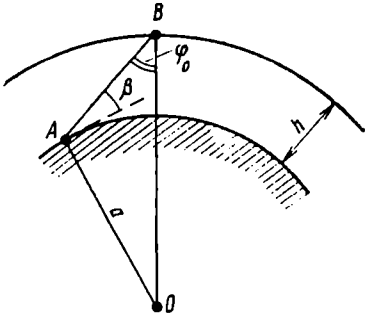


Fig. 4.22. Relationship between angles of elevation and incidence

fixed*, the ray can be made to return to earth by suitably arranging its angle of incidence to the ionosphere.

However, because of the spherical earth, there is an upper limit to the angle of incidence that can be used. As follows from Fig. 4.21, even for a ray tangent to earth's surface (at point A) the maximum angle of incidence is given by

$$\sin \varphi_{0\,max} = a/(a+h)$$

Because of this limit, not all radio waves can be reflected from the ionosphere—some will pass through it under certain conditions.

Referring to Fig. 4.22, let us derive the relationship between the angle of elevation β and the angle of incidence φ_0 to the ionosphere at a height h, that would allow for the effect of the earth's curvature.

From the triangle OAB we have

$$\sin \varphi_0 = \cos \beta/(1+h/a)$$

Substituting this in (4.93), squaring both sides, and neglecting second-order infinitesimals, we get

$$\frac{\cos^2 \beta}{1+2h/a} \cong 1 - 80.8N_n/f^2$$

* Speaking of reflection from a maximum density region, it must be remembered that actually the point of return will lie a little below the peak in all cases.

After simple manipulations, we finally arrive at

$$f = \sqrt{\frac{80.8N_n(1+2h/a)}{\sin^2\beta + 2h/a}} \text{ Hz} \tag{4.101}$$

Equation (4.101) shows that for a fixed angle of elevation, β, a greater electron density is required to reflect a higher frequency. If the maximum electron density is fixed, there will be a limiting frequency defined as

$$f_{max} = \sqrt{\frac{80.8N_{max}(1+2h/a)}{\sin^2\beta + 2h/a}} \text{ Hz} \tag{4.102}$$

known as the maximum usable frequency.

On the other hand, if the angle of elevation is varying, f_{max} will be a maximum with $\beta = 0$, that is, with the rays whose path is horizontal to the ionospheric layer

$$f_{max\,0} = \sqrt{\frac{80.8N_{max}(a+2h)}{2h}} \text{ Hz} \tag{4.103}$$

When $\beta = \pi/2$, that is, for a ray radiated vertically upwards, Eq. (4.102) takes the form

$$f_{max} = f_c = \sqrt{80.8N_{max}} \text{ Hz} \tag{4.104}$$

where f_c is the critical frequency. This is a maximum frequency for which a wave radiated vertically upwards can be returned to earth by the ionosphere.

From comparison of (4.103) and (4.104) it follows that the maximum usable frequency is $\left(\frac{a+2h}{2h}\right)^{1/2}$ times the critical frequency. At $h = 200$ kilometres, the ratio $f_{max\,0}/f_c$ is 4.1.

From (4.102) it follows that a wave will be reflected if and only if the frequency of the wave does not exceed $f_{max\,0}$. Otherwise, the wave will pass through the ionosphere.

Reflection from the Ionosphere at Vertical Incidence. At normal incidence, the ionosphere will reflect only the waves whose frequency does not exceed the critical value given by (4.104) which can be derived directly from (4.93). Thus, if we set $\varphi_0 = 0$ in (4.93), we find that $n = 0$ or, which is the same thing, $\varepsilon_i' = 0$. Thus, a vertically incident wave will be reflected from an ionospheric region where the permittivity reduces to zero. Assuming that at high altitudes the electron number density continues to increase, we conclude that the permittivity at those altitudes grows negative and the refractive index, imaginary.

In § 1.5 it has been shown that in an imperfectly conducting medium radio waves are propagated at a velocity $u = c/n$ [Eq. (1.28)], and the absorption is $\delta = (\omega/c)p$ [Eq. (1.27)], where n and p are

real roots of the system of equations

$$\left.\begin{aligned} n^2 - p^2 &= \varepsilon_i' \\ np &= 30\lambda\sigma \end{aligned}\right\} \tag{4.105}$$

In our case, collisions between the electrons and molecules are neglected, that is, we assume that $\sigma = 0$. Besides, $\varepsilon_i' < 0$ from the statement of the problem. Under the circumstances, the only real solutions for (4.105) are

$$n = 0; \quad p = \sqrt{-\varepsilon_i'} \tag{4.106}$$

where p is real because $\varepsilon_i' < 0$.

Substituting these expressions in (1.26) gives

$$E_z = E_m e^{i\omega\left(t - \frac{xn}{c}\right)} e^{-\frac{\omega}{c} px} = E_m e^{-\frac{\omega}{c}\sqrt{-\varepsilon_i'}x} e^{i\omega t} \text{ V/m} \tag{4.107}$$

The expression derived describes a non-wavelike electric field which rapidly decreases with height, and shows that no waves penetrate beyond the level where $n = 0$.

It we wish to allow for collisions of electrons with neutral molecules and ions, we shall have to treat the ionized gas as an imperfectly conducting medium of permittivity described by (4.54) and conductivity described by (4.55). Substituting these quantities in (1.31) and (1.32) yields the following expressions for the refractive index and the absorption coefficient:

$$n = \pm \sqrt{\frac{1}{2}\left[\varepsilon_i' + \sqrt{\varepsilon_i'^2 + (60\lambda\sigma_i)^2}\right]} \tag{4.108}$$

$$\delta = \pm \frac{2\pi}{\lambda} \sqrt{\frac{1}{2}\left[-\varepsilon_i' + \sqrt{\varepsilon_i'^2 + (60\lambda\sigma_i)^2}\right]} \text{ S/m} \tag{4.109}$$

It is essential to note that with the collisions of electrons allowed for the refractive index will not reduce to zero even with negative values of permittivity.

The Effect of the Earth Magnetic Field on the Reflection of Radio Waves. To begin with, consider the effect of the terrestrial field on the reflection of waves at vertical incidence.

Since we ignore collisions with molecules, it is natural to assume that the conditions for a wave to be returned to earth will remain the same, that is, $n_1 = 0$ for the extraordinary ray and $n_2 = 0$ for the ordinary ray.

Referring to (4.72), we find that for the ordinary ray n_2 reduces to zero when $\omega = \omega_0$. Noting (4.75), we find that this corresponds to the reflection from the ionosphere in the absence of a magnetic field.

The conditions under which the extraordinary ray will be returned to earth can be ascertained by equating the denominator of (4.72) to ω_0^2

$$\omega^2 - \frac{\omega^2\omega_t^2}{2(\omega^2-\omega_0^2)} - \sqrt{\left[\frac{\omega^2\omega_t^2}{2(\omega^2-\omega_0^2)}\right]^2 + \omega^2\omega_l^2} = \omega_0^2$$

or, after simplifications,

$$\omega^2 - \omega_0^2 = \pm\omega\omega_m \tag{4.110}$$

where

$$\omega_m^2 = \omega_l^2 + \omega_t^2 = \left(\frac{\mu_0 e H_0}{m}\right)^2 \text{ s}^{-2}$$

Solving for ω, we find that for the extraordinary ray to be reflected the condition is given by

$$\omega', \omega'' = \omega_0\sqrt{1 + \left(\frac{\omega_m}{2\omega_0}\right)^2} \mp \frac{\omega_m}{2} \tag{4.111}$$

From (4.111) it follows that $\omega' < \omega_0$, and $\omega'' > \omega_0$.

It is important that the condition for the reflection of the extraordinary ray is independent of the relative magnitude of the magnetic-field components and is solely decided by the total field of the earth

$$H_0 = \sqrt{H_l^2 + H_t^2} \text{ A/m}$$

To identify which of the frequencies, ω' or ω'', corresponds to the real reflections of the extraordinary ray, let us refer to Fig. 4.23 which is a plot of the refractive index squared as a function of the ratio $(\omega_0/\omega)^2$, in accordance with (4.72). For a given frequency, this ratio is proportional to the electron density.

In plotting the relationship, it has been assumed that $\omega = 2\omega_m$ (which corresponds to a wavelength of about 100 metres) and $\alpha = 45°$. As may be seen, the relationship for the extraordinary ray is more elaborate. When $\omega_0/\omega = 1$, the function $n^2 = F(\omega_0/\omega)$ shows a discontinuity. From the plot of Fig. 4.23 it is also seen that as the electron density increases, the refractive index decreases for both components, reaching zero at point A for the extraordinary ray (when $\omega_0/\omega < 1$) and at point B for the ordinary ray (when $\omega_0/\omega = 1$). This indicates that under actual conditions the wave of frequency ω'' will be reflected. Thus, in the short-wave band, when $\omega > \omega_m$, the frequency of the extraordinary ray reflected from a region with a given electron density exceeds the frequency of the ordinary ray. For a fixed frequency, the extraordinary ray will be reflected from a higher level than the ordinary one.

When the ray is nearly horizontal to the layer, the condition for the ray to return to earth is

$$\sin^2\varphi_0 = n_2^2 \tag{4.112}$$

for the ordinary ray, and

$$\sin^2 \varphi_0 = n_1^2 \qquad (4.112a)$$

for the extraordinary ray, where n_1 and n_2 are the refractive indexes for the extraordinary and ordinary ray, respectively, as given by (4.72). Solving these equations for frequency, we can get expressions for the maximum frequency analogous to (4.102) for the extraordinary and ordinary rays.

At vertical incidence, the ordinary and extraordinary components are propagated along a common path, except the segment between

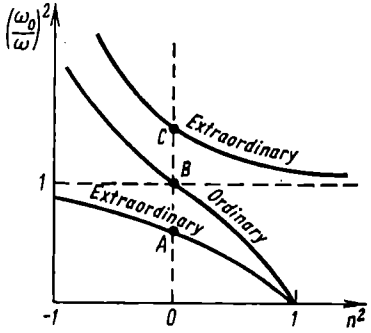


Fig. 4.23. The square of the refractive index as a function of the relative frequency

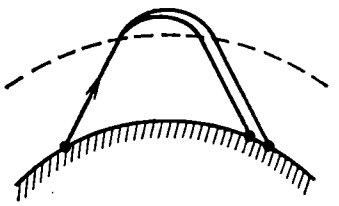


Fig. 4.24. Splitting of a radio wave by the earth's magnetic field

the reflection points of the two components. With oblique incidence, the difference in the refractive indexes results in double refraction, so that the ordinary and extraordinary components take different paths, as shown diagrammatically in Fig. 4.24.

A more detailed discussion of refraction from the ionosphere will be found in the chapters dealing with the propagation of sky (ionospheric) waves.

4.11. Ionospheric Stations

Until quite recently, ground-based ionospheric stations were the main tool of studies into the ionized layers of the earth's atmosphere. It is only for some time past that direct measurements of the basic ionospheric indices have become possible through the instruments carried aloft by geophysical rockets, earth satellites, and space probes. Yet, these direct measurements cannot fully supplant the observations carried by ionospheric stations throughout the planet on the round-the-clock basis.

As regards the ionosphere, present-day ionospheric stations are doing the same work as the world-wide network of weather stations does for the lower regions of the atmosphere. While the latter furnish information for the weather service, the former supply it for the "ionosphere prediction service". Space flights also yield valuable information which goes to supplement the data gathered by ionospheric stations daily.

At present, there exist three types of ionospheric stations—vertical-incidence sounding, oblique-incidence sounding, and oblique incidence-backscatter sounding.

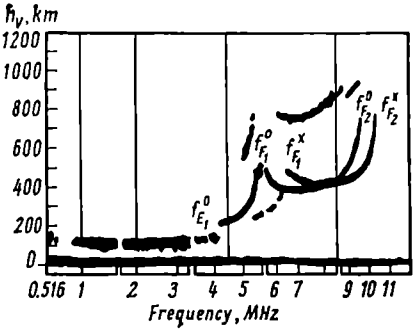


Fig. 4.25. An example of an ionogram

A vertical-incidence sounding station is a combination of a transmitter and a receiver placed side by side and often using the same aerial. The transmitter sends out pulses of electromagnetic energy, so that during an observation (usually lasting for about a minute) its frequency is swept through a range from 0.5 to 20 megahertz. At the same time, the receiver is automatically held on tune to the swept frequency of the transmitter. The output of the receiver is coupled to a display (usually, a cathode-ray tube) which makes it possible to measure and record (on a photograph or a cine film) the time delay between the transmitted and returned pulse as a function of the swept frequency of the transmitter. Assuming that the waves travel to and back from the reflection point at the same velocity as does light in free space, the time delay τ can be used to determine the virtual height of the ionized layers

$$h_v = 3 \times 10^8 \tau/2 \text{ m} \qquad (4.113)$$

This height is called "virtual" because Eq. (4.113) ignores the fact that in an ionized atmosphere radio pulses travel at the group velocity $u_g < c$. A plot of the function $h_v = F(f)$ is called an ionogram or a height-vs-frequency characteristic. Examples of ionograms appear in Figs. 4.25 and 4.26 (the latter is presented in sketch form). Using

ionograms, one can readily determine the heights of reflecting layers and, what is especially important, the critical frequencies reflected from the individual ionized regions. The respective critical frequencies bear the designations $f_c\,(E)$, $f_c\,(F_1)$, etc., in Fig. 4.26.

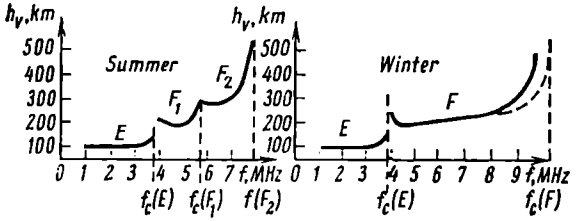


Fig. 4.26. Height-vs-frequency characteristics of the ionosphere

Two facts stand out when one looks at an ionogram. Firstly, the virtual heights steeply increase as the critical frequencies are approached. As will be shown shortly, this is an apparent increase due to

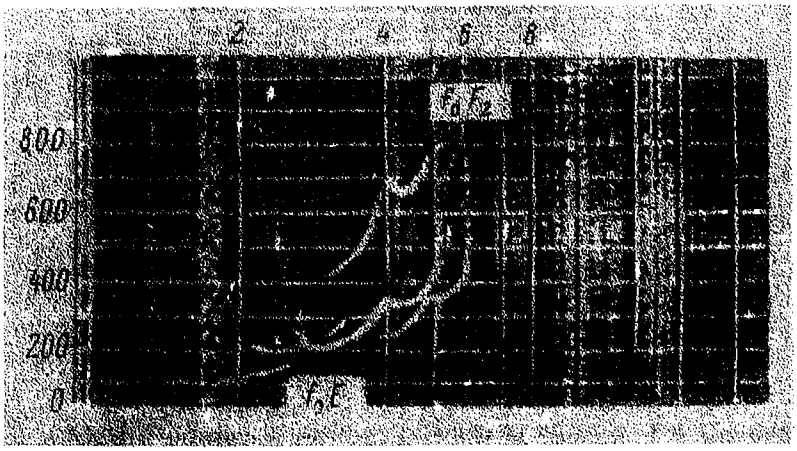


Fig. 4.27. Panoramic ionogram

the group delay. Secondly, the waves reflected from the *F1* and *F2* layers, which are of an unquestionably electron constitution, experience double refraction. The ordinary and extraordinary rays have a different critical frequency which is denoted f^{o}_{cF1} for the ordinary ray and f^{x}_{cF1} for the extraordinary ray.

Present-day ionospheric stations are highly automated outfits. The more sophisticated of them are panoramic stations pioneered

by Bulatov of the Soviet Union. The display is a cathode-ray tube producing a complete ionogram which can be photographed at specified time intervals on cine film. An example of the ionogram is shown in Fig. 4.27.

Let us derive the relation between the virtual and actual heights of reflection points. In Fig. 4.28,*h* stands for the height at the entrance to the ionosphere. Radio waves cover this path with the velocity of light in free space. The height of the reflection point above the lower edge of the ionosphere is denoted z_0, and the virtual height of reflections, h_v.

Denoting the electron density as a function of height as $N(z)$, the time delay of the reflected pulse relative to the exploring pulse may be given by

$$\tau = \frac{2h_v}{c} = \frac{2h}{c} + 2\int_0^{z_0} dz/u_g \text{ s} \qquad (4.114)$$

The integral in (4.114) gives the group delay.

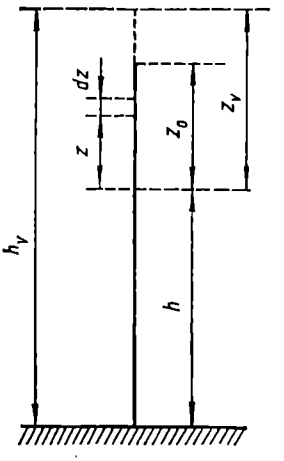


Fig. 4.28. Relationship between virtual and actual heights

Substituting $u_g = cn$, where $n = \sqrt{1 - 80.8N(z)/f^2}$, and cancelling, we get

$$h_v = h + \int_0^{z_0} \frac{dz}{\sqrt{1 - 80.8N(z)/f^2}} \text{ m} \qquad (4.115)$$

It is to be noted that the denominator of the integrand is less than unity, that is, the integral is always greater than unity. In other words, the virtual height h_v is always greater than $(h + z_0)$. This is not unexpected, because in defining the virtual height we assumed that the total path was travelled at a greater velocity than this is actually the case.

In § 4.10 it is shown that for a ray at vertical incidence to be brought back to earth it is essential that $\varepsilon_i' = 1 - 80.8N(z)/f^2 = 0$. Than is, when $z = z_0$, the denominator of the integrand reduces to zero. As will be recalled, such integrals are called improper. According to Cauchy's criterion, this type of improper integrals is convergent (that is, has a finite value) if for z close to z_0 the following condition is satisfied

$$|f(z)| < \frac{C}{(z_0 - z)^p} \qquad (4.116)$$

where $f(z)$ is the integrand, and C and p are positive constants such that $p < 1$.

In our case, we seek to determine the condition under which the product $|f(z)|(z_0 - z)^p$ remains bounded for $z \to z_0$.

On setting $p = 0.5$, we get

$$\lim_{z \to z_0} [\,|f(z)|(z_0 - z)^{0.5}\,] = \lim_{z \to z_0} \sqrt{\frac{z_0 - z}{1 - 80.8N(z)/f^2}}$$

$$\cong \frac{f}{9\sqrt{(dN/dz)_{z=z_0}}}\, m^{1/2} \qquad (4.117)$$

From (4.117) it follows that the improper integral ceases to be convergent only in the region of the ionosphere where the vertical gradient of the electron density reduces to zero, that is, where the density is a maximum. This is why ionograms show a sudden increase in the virtual heights closer to the critical frequencies.

At an oblique-incidence sounding station, the transmitter and receiver are located at the end points of the propagation path. The indicator (a cathode-ray tube) of the receiver is synchronized by the exploring pulses sent out by the transmitter which is swept in frequency through a specified frequency band. The ionogram thus obtained relates the time delay to the swept frequency of the transmitter (the time delay can be readily converted to the distance covered by the exploring wave). Sometimes, it is called a distance-frequency characteristic. Using it, one can predict optimum communication frequencies (in the short-wave band) for the link served by a given pair of ionospheric stations.

Because of the difficulty of synchronizing the receiver and transmitter, oblique-incidence sounding stations have for the most part been replaced with oblique incidence backscatter sounding stations. In such a station, the transmitter and receiver are at the same location, as they are in a vertical-incidence sounding station. However, the aerials (similar to those used by short-wave transmitters) radiate electromagnetic energy at a low angle to the horizon. Instead of the waves reflected from the ionosphere, the receiver registers those scattered from the earth's surface where the steepest wave reflected from the ionosphere strikes it (as shown in Fig. 4.29). The scattering of radio energy from the earth's surface and return of some of the scattered energy to the station after a second reflection from the ionosphere is known as the Kabanov effect after Kabanov of the Soviet Union who studied the phenomenon in 1947-1958 [77].

Incidentally, Kabanov has shown that practically all types of surface are rough, although to different degrees, and so a scattered signal can be picked up in each case.

Oblique-incidence backscatter sounding stations operate on the same principle as oblique-incidence sounding stations, but the

transmitter and receiver are placed side by side so that there is no need for synchronizing the indicator. By use of rotatable aerials and P-type (plan-position) display, it is possible to obtain distance-frequency characteristics for any direction of propagation in a horizontal plane. Another distinction is that while an oblique-incidence sounding station can provide a check on radio propagation along

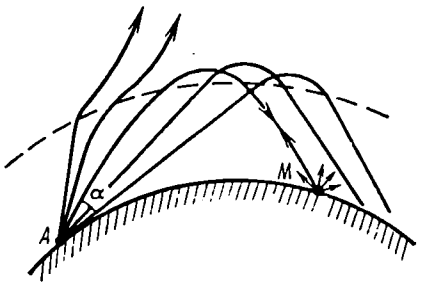


Fig. 4.29. Ray paths in oblique-incidence backscatter sounding

a specific path, an oblique-incidence backscatter sounding station is more versatile since it can be used on a particular path in conjunction with a fixed aerial and on any direction in conjunction with a rotatable aerial.

Referring to Fig. 4.29, the rays at vertical incidence pass through the ionosphere without being reflected from it. The steepest of all ionosphere-reflected rays is incident at point M on the earth's surface, and some of the scattered energy retraces the same path. The more oblique rays strike the earth's surface farther away from the station and are also scattered. The ionogram clearly shows the rays scattered from the nearest point.

The distance-frequency characteristics obtained by oblique-incidence backscatter sounding stations have the appearance sketched in Fig. 4.30. In the interval from f_1 to the maximum usable frequency (MUF) reflections occur from two heights in the ionosphere, because of which there is a sort of loop on the ionogram.

4.12. Virtual Ionospheric Height and Path

While oblique-incidence and oblique-incidence backscatter sounding stations can directly determine the maximum usable frequency (in the short-wave band) vertical-incidence sounding stations can only determine the conditions for waves at vertical incidence to be brought back to the ground. This entails the problem of re-calculating vertical-incidence sounding data into those evaluating the propagation of the waves employed in a practical radio circuit.

It may appear that the simplest way to do this is by means of the secant law (4.100)

$$f = f_{vert} \sec \varphi_0$$

which relates a vertical and an oblique ray reflected from the same layer. However, to be able to use the secant law, one must know the angle of incidence to the ionosphere. This angle can be found by the two theorems set forth below.

The first theorem relates the virtual heights of an oblique and a vertical ray reflected at the same real height z_0. Let the virtual

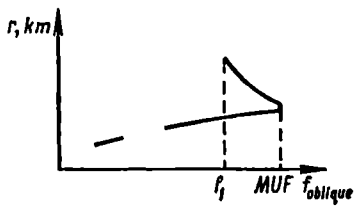


Fig. 4.30. Distance-vs-frequency characteristic

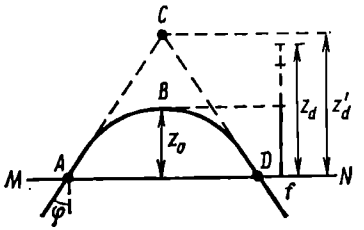


Fig. 4.31. Virtual ionospheric height and path

height z'_v, of the oblique ray be the height of the corner C of the triangle (Fig. 4.31) obtained by extending the straight-line propagation paths of radio waves in air until they intersect. Denoting the distance AD as r, from Fig. 4.31 we have

$$z'_v = \frac{1}{2} r \cot \varphi_0 = \frac{\cot \varphi_0}{2} \int_s \sin \varphi \, ds \text{ m} \qquad (4.118)$$

where ds = element of the curved path of the wave;
φ = variable angle that the path element makes with the vertical.

The subscript s indicates that the integral is taken over the entire curved path between the points A and D.

Placing the term $1/\sin \varphi_0$ under the integral, assuming that $\sin \varphi_0 = n \sin \varphi$ at each point of the path according to the law of refraction, and, finally, passing from the variable s to the new variable z under the sign of integral, we get

$$z'_v = \frac{\cos \varphi_0}{2} \int_s \frac{ds}{n} = \cos \varphi_0 \int_0^{z_0} \frac{dz}{n \cos \varphi} \text{ m} \qquad (4.119)$$

Using once more the law of refraction which holds for any point of the path, and writing n in terms of the electron density, the deno-

minator of the integrand may be given the form

$$n\cos\varphi = \sqrt{n^2 - n^2\sin^2\varphi} = \sqrt{n^2 - \sin^2\varphi_0} = \sqrt{\cos^2\varphi_0 - 80.8N(z)/f^2}$$

Writing f in terms of f_{vert} by use of the cosecant law, we get

$$n\cos\varphi = \cos\varphi_0\sqrt{1 - 80.8N(z)/f_{vert}^2}$$

Taking this expression as the denominator of the integrand in (4.119), we arrive at the following expression for the virtual height of the oblique ray

$$z_v' = \int_0^{z_0} \frac{dz}{\sqrt{1 - 80.8N(z)/f_{vert}^2}} \text{ m} \tag{4.120}$$

which is exactly the same as (4.115) for the virtual height of the vertical ray. Hence, we may write

$$z_v' = z_v \text{ m} \tag{4.121}$$

and state the first theorem as follows:

If a vertical ray of frequency f_{vert} and an oblique ray of frequency f are reflected from one and the same region in the ionosphere, that is, if the reflection points are at the same real height, the virtual height is the same for both rays.

The second theorem relates the time of travel for a signal along the real and the virtual paths. The time T during which the signal travels along the real curved path in the ionosphere is given by

$$T = \int_s ds/u_g = 2\int_0^{z_0} \frac{dz}{cn\cos\varphi} = \frac{2}{c}\int_0^{z_0} \frac{dz}{n\cos\varphi} \text{ s} \tag{4.122}$$

On comparing (4.122) and (4.119), we may write

$$T = \frac{2z_v'}{c\cos\varphi_0} \text{ s} \tag{4.123}$$

On the other hand, referring to Fig. 4.31 and noting that $AC = z_v/\cos\varphi_0$, the time of travel along the virtual triangular path at the velocity of light in free space is

$$T_v = 2z_v'/c\cos\varphi_0 \text{ s} \tag{4.124}$$

which precisely checks with the earlier expression for T. Thus

$$T = T_v \text{ s} \tag{4.125}$$

Eq. (4.125) conveys the gist of the second theorem which may be stated as follows:

The time of travel of a radio wave around the real curved path in the ionosphere is equal to the time of travel around the virtual triangular path at the velocity of light in free space.

The value of the two theorems is that the real curved path for a wave of frequency f may be replaced with the virtual path of a ray propagated at the velocity of light in free space. The virtual path is found by constructing an isosceles triangle with AB as the base (Fig. 4.32), where A and B are the terminals of the radio link, with the height $h + z_v$, which is the virtual height of reflection for the

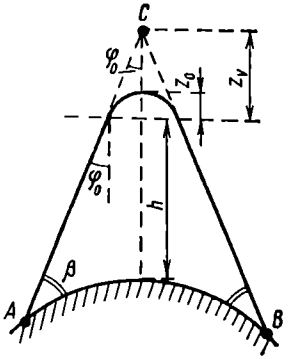


Fig. 4.32. Actual path (the full line) and virtual path (the broken line) of a radio wave

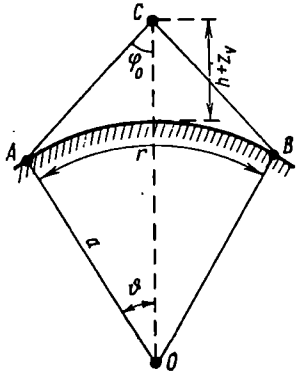


Fig. 4.33. Relationship between the angle of incidence at the entrance to the ionosphere and the virtual height of a reflection point

equivalent frequency at normal incidence as found by ionospheric measurements and related to the frequency of the oblique ray by the secant law

$$f_{vert} = f/\sec \varphi_0 \text{ Hz}$$

Referring to Fig. 4.33 where r is the propagation path length and θ is the central angle corresponding to a half of the arc r, from the triangle OAC we have

$$\frac{\sin \varphi_0}{a} = \frac{\sin (180° - \varphi_0 - 0)}{a + h + z'_v} \text{ m}^{-1}$$

Solving for φ_0 gives:

$$\tan \varphi_0 = \frac{a \sin 0}{h + z'_v + a (1 - \cos \theta)} \tag{4.126}$$

On the other hand, setting $a = 6370$ kilometres, we get

$$\theta° = r_{(\text{km})}/222.4 \tag{4.127}$$

Thus, the "virtual-path" approach markedly simplifies calculations and, among other things, makes it possible to determine the angle of incidence at the entrance to the ionosphere.

4.13. Ionospheric Maps

Ionospheric propagation (which is discussed at length in the next chapter) is mainly determined by the condition of the ionosphere. Watch on its condition is continually held by a world-wide network of ionospheric stations, now operating even in the Arctic and Antarctic. Their findings are presented on so-called ionospheric maps compiled for each of the basic regions and layers of the ionosphere (*D*, *E*, *F1*, and *F2*).

From § 4.6 it follows that the *D* and *E* regions and the *F1* layer are regular features of the ionosphere. Their ionization is decided almost single-valuedly by the zenith angle of the sun and, to some extent, by solar activity. Therefore, their maps may be compiled once and for all, and there is no need to update them continually as weather maps are. In contrast, the *F2* layer, being the outer layer of the ionosphere, is exposed more to solar radiant and corpuscular emissions, with the result that its characteristics vary suddenly from day to day and with changes in solar activity and geomagnetic disturbances. Therefore, there is the need for maps of the *F2* layer showing what is to be expected in several months to come, similar to those used in long-term weather forecasts. As a matter of record, ionospheric predictions are usually more accurate than weather forecasts.

It is customary to compile two types of ionospheric map for the *F2* layer, such as shown in Figs. 4.34 and 4.35 [78]. For simplicity, such maps are constructed on a rectangular equidistant projection. Longitude is laid off as abscissa, and latitude as ordinate. Soviet ionospheric maps describe the state of the ionosphere at a specified hour of Moscow decree time. As a rule, maps are compiled at two-hour intervals, so that a daily complement consists of twelve maps. Of course, the continents on such maps are shown with their contours greatly distorted, but this is of minor consequence to radio calculations. The map of Fig. 4.34 shows isopleths, curves connecting locations with the same value of the critical frequency, that is, the maximum frequency that can be reflected from a given ionospheric layer at a given electron density and vertical incidence. The numerals at the isopleths give these critical frequencies in megahertz. Thus, from the map of Fig. 4.34 we can determine the expected critical frequency at any location all over the world at a specified hour (Moscow time).

Unfortunately, a map showing isopleths of critical frequency does not—and cannot, for that matter,—give any idea about the virtual heights of a reflecting region. On the other hand, exactly this quantity is important in the determination of maximum frequencies, because, as we have seen, Eq. (4.102) for f_{max} includes the height

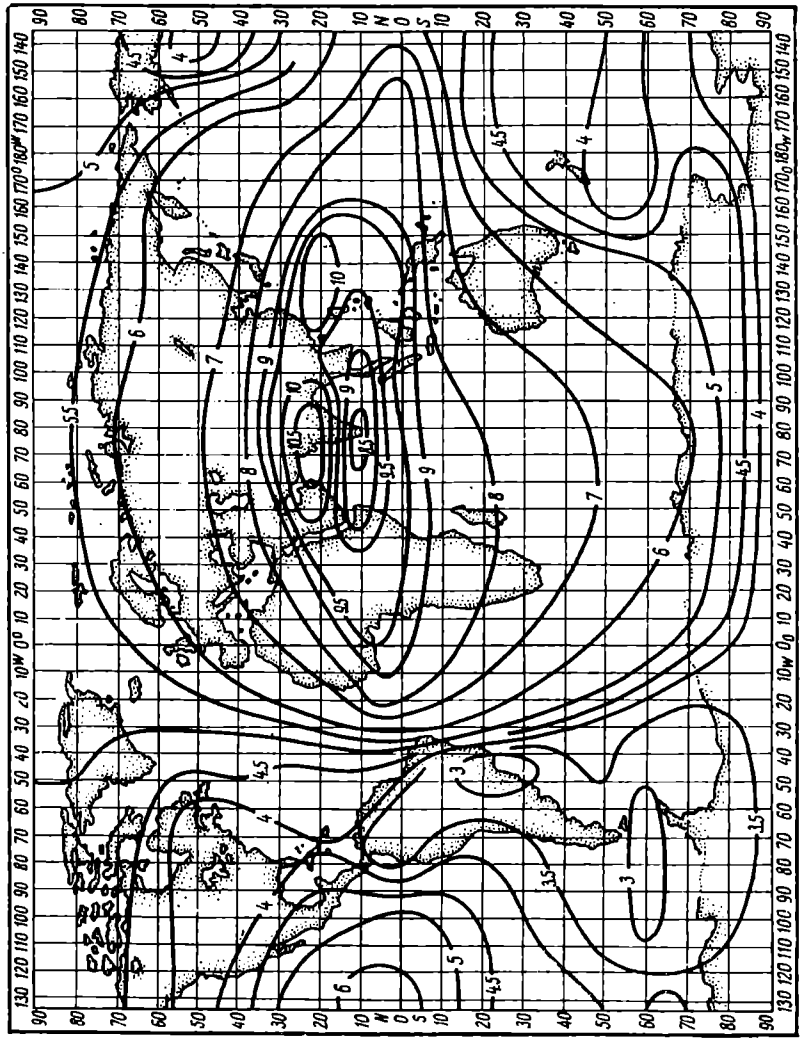


Fig. 4.34. *F*2 ionospheric map for MUF-0 (forecast for August, 1962)

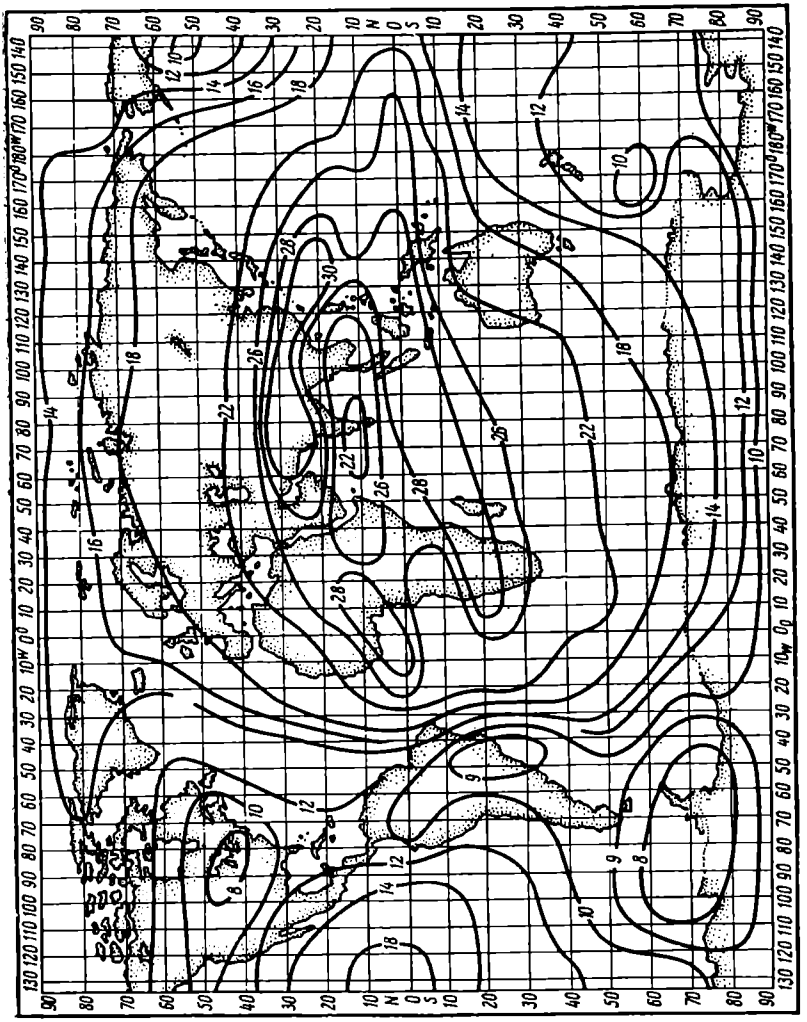


Fig. 4.35. *F*2 ionospheric map for MUF-4000 (forecast for August, 1962)

of the reflecting region. This is the reason why a second type of map, such as shown in Fig. 4.35, is compiled. Such maps show contours of maximum usable frequencies (MUFs) for zero and 4000-km distance between the transmitter and the receiver. The map of Fig. 4.35 has been compiled for MUF-4000, that is, for the transmitter and receiver separated by a distance of 4000 kilometres. The numerals on the isopleths give the maximum usable frequency in megahertz. A given MUF applies to a particular radio circuit, if the geographical coordinates of the middle point of the circuit lie on the respective isopleth.

Ionospheric maps reflect regular features of the ionosphere, namely, the geographical location, the time of the day, the season of the year, and the phases of the 11-year solar cycle. These factors are all combined to predict monthly median parameters of the *F2* layer. Unfortunately, such maps cannot predict changes in the state of the *F2* layer that may be caused by ionospheric disturbances and other irregular events.

There are separate maps for the Sporadic *E* layer. Instead of the critical frequency of the layer, such a map shows the probability of occurrence of the *Es* layer whose critical frequency exceeds a specified value (usually 3 megahertz). Using a suitable nomogram and the known probability of occurrence of such a layer, one can readily determine the probability of occurrence of the Sporadic *E* layer with other values of the critical frequency.

Chapter Five

Ionospheric Propagation

5.1. Physical Processes in the Propagation of Sub-audio Waves. Magnetohydrodynamic Waves

According to the classification adopted in this text, the term sub-audio will refer to radio waves lying in the frequency range from 3 millihertz to 10 hertz, and the term audio to radio waves occupying the frequency interval from 10 hertz to 3 kilohertz, so that their high-frequency end is directly adjacent to very long waves (or very low frequencies). So far, these wavelengths have not yet found any practical uses, and their origin lies in natural phenomena — movement of the ionized air, lighting discharges, etc.

Present-day physics treats the ionized atmosphere as a fourth state of matter, that is, as a rarefied plasma placed in the terrestrial magnetic field. As Alfvén [79] discovered in 1942, the electric current induced by the motion of an ionized gas across a magnetic field interacts with the magnetic field and gives rise to a wave-like motion of the ionized gas. These waves, called *magnetohydrodynamic* by Alfvén, have been found to travel along magnetic field lines. Since this mode is inseparably linked with the electromagnetic field, some authors call these waves electromagnetohydrodynamic. Although more precise, this term is unwieldy, and we shall use magnetohydrodynamic (or, rather, MHD) instead.

Working on Alfvén's idea, Ginzburg [68] has shown that MHD waves are the limiting case of sub-audio waves propagated in the ionosphere along the magnetic-field lines.

Let there be a radio wave of a very low frequency ω travelling in the electron-iron plasma placed in the earth's permanent magnetic field. We assume that the number of free electrons per unit volume, N, is equal to that of positive ions, N_+.

For waves propagated in a longitudinal magnetic field the refractive index is

$$n_{1,2}^2 = 1 - \frac{\omega_0^2}{\omega(\omega \mp \omega_l)} \tag{5.1}$$

Substituting the expression for ω_0^2 from (4.75) in (5.1) gives

$$n_{1,2}^2 = 1 - \frac{Ne^2}{\omega\varepsilon_0 m(\omega \mp \omega_l)} \tag{5.1a}$$

However, this equation accounts only for electrons of mass m, while the plasma of our example also contains ions. To allow for their contribution, we use the additive law (because the medium is linear) and add the respective term to the right-hand side of the equation. We used the same technique in deriving Eq. (4.56) in § 4.7. Now we must allow for the effect of the terrestrial magnetic field on the motion of ions because of which the aggregate effect of ions and electrons will result in the following expression for the refractive index

$$n_{1,2}^2 = 1 - \frac{Ne^2}{\varepsilon_0 m\omega(\omega \mp \omega_l)} - \frac{Ne^2}{\varepsilon_0 m_l \omega(\omega \mp \Omega_l)} \tag{5.2}$$

where m_l is the mass of an ion which for atomic oxygen is 25,800 times that of an electron, while ω_l and Ω_l are the gyro-frequencies of electrons and ions defined by Eqs. (4.75) as

$$\omega_l = \mu_0 eH_0/m \text{ s}^{-1} \tag{5.3a}$$

$$\Omega_l = \mu_0 eH_0/m_l \text{ s}^{-1} \tag{5.3b}$$

Setting the strength of the terrestrial magnetic field at $H_0 = 40$ amperes per metre, the gyro-frequency for electrons is $f_l = 1.4$ megahertz. For ions of atomic oxygen the gyro-frequency is

$$\Omega_l = \frac{1.4 \times 10^6}{2.58 \times 10^4} = 54 \text{ Hz}$$

Our further discussion will hold for frequencies

$$\omega \ll \Omega_l \text{ s}^{-1} \tag{5.4}$$

that is, we shall be dealing with frequencies from one hertz down.

It may be added that in the absence of the terrestrial magnetic field the second term in (5.2) would be 1/25,800th of the first and it might be safely ignored, which was done in § 4.8.

It is also to be noted that electrons and ions carry unlike charges, and so they rotate in opposite directions. This is allowed for by reversing the signs in the last term of (5.2).

After simple manipulations, (5.2) may be re-written as

$$\begin{aligned} n_{1,2}^2 &= 1 - \frac{Ne^2}{\varepsilon_0 \omega m}\left(\frac{1}{\omega \mp \omega_l} + \frac{m/m_i}{\omega \pm \Omega_l}\right) \\ &= 1 - \frac{Ne^2}{\varepsilon_0 \omega m}\left(\frac{\omega \pm \Omega_l + (m/m_i)\,\omega \mp \omega_l m/m_i}{\omega^2 \pm \omega\Omega_l \mp \omega_l\omega - \omega_l\Omega_l}\right) \end{aligned} \tag{5.2a}$$

In the numerator, the sum $(\pm\Omega_l \mp \omega_l\, m/m_i)$ is identically equal to zero. Besides, the third term may be neglected in comparison

with the first, because $m/m_i = 1/25{,}800$. Owing to (5.4) and also because $\omega_l \gg \omega$, we may neglect the first three terms in the denominator of (5.2a) in comparison with the last. Finally, we get

$$n^2 = 1 + Ne^2/\varepsilon_0 m\omega_l\Omega_l \tag{5.2b}$$

Substituting the expressions for the gyro-frequencies from (5.3a) and (5.3b) in (5.2b) and noting that the second term in (5.2b) is many times greater than unity, we have

$$n^2 \cong Nm_i/\varepsilon_0\mu_0^2H_0^2 = \rho/\varepsilon_0\mu_0^2H_0^2 \tag{5.5}$$

where $\rho = Nm_i$ is the density of the plasma in kilograms per cubic metre, thus expressed on the assumption that the mass of electrons may be neglected in comparison with that of ions.

The velocity of MHD wave propagation is given by

$$u = c/n = \frac{c\sqrt{\varepsilon_0\mu_0}\sqrt{\mu_0}H_0}{\sqrt{\rho}} = H_0\sqrt{\mu_0/\rho} \text{ m/s} \tag{5.6}$$

It may be added in passing that at sufficiently low frequencies the refractive index is independent of frequency. Exactly the same expression for the velocity of propagation was derived by Alfvén [19], although in a different, hydrodynamic manner.

The example that follows shows that the absolute velocity of MHD waves in the ionosphere is relatively low.

Example 5.1. Given: $\rho = 10^{-15}$ g/cm^3 $= 10^{-12}$ kg/m^3; $H_0 =$ 40 A/m. To find: the velocity of MHD waves at sub-audio frequencies.

Solution: Substituting the above values and $\mu_0 = 4 \times 10^{-7}$ henry per metre in (5.6) gives

$$u = 40\sqrt{\frac{4\pi \times 10^{-7}}{10^{-12}}} = 4.5 \times 10^4 \text{ m/s}$$

By way of comparison, the velocity of sound in the plasma is

$$u_s \cong \sqrt{kT/m_i} \text{ m/s} \tag{5.7}$$

where $k = 1.38 \times 10^{-23}$ joule per degree. Substituting the mass of an oxygen ion $m_i = 9.11 \times 10^{-31} \times 25{,}800 = 2.36 \times 10^{-26}$ kilogram and assuming that $T = 1000$ °K, we get

$$u_s = 760 \text{ m/s}$$

The role that MHD waves play in the earth's ionosphere has not yet been fully elucidated. It may be presumed that MHD waves control fluctuations in the ionosphere that result in fading in the case of ionospheric propagation. It may be likewise presumed that MHD waves figure prominently in astrophysics.

5.2. Physical Processes in the Propagation of Audio-frequency Waves. Whistling Atmospherics

While MHD waves, that is, waves at sub-audio frequencies, are generated by the motion of the plasma, radio waves at audio-frequencies have their origin in lighting flashes occurring in the troposphere, that is, below the ionized region of the earth's atmosphere. Such discharges with a current of up to 100,000 amperes and a duration of about 100 microseconds generate radio waves over a fairly broad frequency spectrum covering part of the audio band.

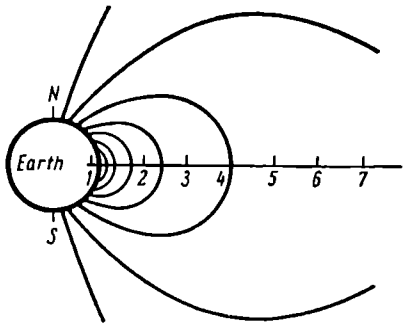


Fig. 5.1. Lines of the terrestrial magnetic field along which audio-frequency radio waves are propagated

So far we have assumed that the waves at frequencies below the critical ones (see Eq. 4.102 in § 4.10) are always reflected from the ionosphere, that is, they are unable to enter and propagate in the ionosphere. However there is an exclusion to this rule.

If we allow for the effect of the terrestrial magnetic field and assume that radio waves travel along the magnetic-field lines then the refractive index for the extraordinary wave at a sufficiently low frequency will not turn to zero, which is an indication that the ray can enter the ionosphere. Fig. 5.1 shows that the lines of the terrestrial magnetic field are at nearly right angles to the earth's surface only in areas close to the magnetic poles. This implies that the audio-frequency waves generated by lighting flashes in the polar areas may well travel through the ionosphere.

For the extraordinary component of a wave propagated in a longitudinal magnetic field the refractive index has been found to be

$$n_1^2 = 1 - \frac{\omega_0^2}{\omega(\omega - \omega_l)} \tag{5.8}$$

The group velocity of radio waves, as was shown in § 4.8, is

$$u_g = c/(n + \omega\, dn_1/d\omega) \text{ m/s} \tag{5.9}$$

Unfortunately, $u_p u_g = c^2$ [Eq. (4.65a)] does not hold for an anisotropic medium, and in order to determine the group velocity one has to make all operations indicated in (5.9). On differentiating, we get

$$u_g = 2c\, \frac{(\omega_l - \omega)^{3/2}\, [\omega^2 (\omega_l - \omega) + \omega\omega_0^2]^{1/2}}{2\omega^3 - 4\omega^2\omega_l + 2\omega\omega_l^2 + \omega_l\omega_0^2} \text{ m/s} \tag{5.10}$$

In all cases of practical interest, the inequality $\omega_0^2 \gg \omega\omega_l$ holds. Besides, noting that $\omega_l \gg \omega$, we may drop the first term in the square brackets and the first three terms in the denominator. After simplifications we have

$$u_g = 2c \frac{\omega^{1/2} (\omega_l - \omega)^{3/2}}{\omega_l \omega_0} \text{ m/s} \tag{5.10a}$$

The group velocity is a maximum at $\omega = \omega_l/4$, and it reduces to zero when $\omega = 0$ and when $\omega = \omega_l$ [80]. When $\omega \ll \omega_l$, (5.10) takes a still simpler form

$$u_g \cong 2c \sqrt{\omega\omega_l}/\omega_0 \text{ m/s} \tag{5.10b}$$

The magnetic-field lines produce a strong ducting effect on audio-frequency radio waves. Owing to this effect, the radio waves travelling along the field lines move into the opposite hemisphere, covering distances up to ten earth radii [81].

Eq. (5.10b) shows that in contrast to MHD waves the group velocity of audio-frequency waves is frequency-dependent: for the most part, the high frequencies of the spectrum travel faster than the low ones. As a result, at the output of the receiver there appears a whistling tone of a steadily falling pitch. Incidentally, this is why this type of naturally occurring transient electromagnetic disturbance has been called *whistlers*.

Whistling atmospherics provide a powerful technique for studying the constitution of the outermost regions of the ionosphere.

5.3. Physical Processes in the Propagation of Very Long and Long Waves

The upper limit for the VLW band (in terms of wavelength) has not yet been set with precision. For the purpose of our discussion we shall assume that it extends from 10^4 to 10^5 metres, while the LW (long-wave) band extends from 10^3 to 10^4 metres. Obviously, no precise demarcation line can be drawn between very long and audio-frequency waves. Therefore, our findings about the propagation of audio-frequency waves apply to a certain extent to the longest of very long waves.

Very long and long waves are propagated like ionospheric waves—through consecutive reflections between the surface of the earth and the lower edge of the D region by day and of the E region at night. It may be said that very long and long waves travel in a sort of spherical waveguide in which the imperfectly conducting surface of the earth acts as the inner wall and the lower edge of the ionosphere as the outer wall. Much as in a metal waveguide the elementary rays reflected from the waveguide walls combine to produce a ducted

flow of energy, so the long waves reflected from the ionosphere and the earth combine to produce the electromagnetic field of the waves ducted by the ionosphere and therefore caused to travel round the spherical surface of the earth.

According to ionospheric measurements, the lower edge of the E region at night (and, in all probability, that of the D region by day) is a sharply bounded limit. At a height of about a wavelength the electron density in these regions rapidly increases. This implies that very long and long waves are reflected from the lower edge

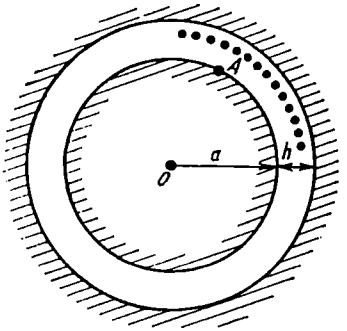


Fig. 5.2. The spherical waveguide in which very long and long radio waves are propagated

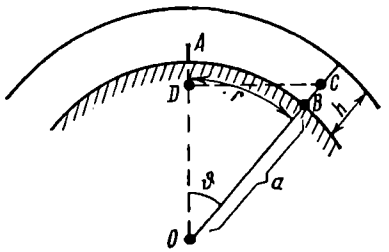


Fig. 5.3. Power flux in the propagation of very long and long waves

of the ionosphere without penetrating very far into these regions. As regards very long and long waves, the D and E regions behave like imperfectly conducting media.

An insight into the propagation of very long and long waves can be gleaned from consideration of the problem which involves the propagation of the waves radiated by an aerial A (Fig. 5.2) inside a spherical waveguide. Apart from the unusual form of the waveguide, the problem is complicated by the fact that both the inner and outer surfaces of the waveguide are imperfect conductors.

As a rough approximation, let us consider the case where the surface of the earth and the lower edge of the ionosphere may be treated as perfect reflectors. Let us further assume that at a sufficiently long distance from the radiator the power flux density becomes constant owing to a series of interactions between the rays undergoing a different number of reflections from the waveguide walls.

Denoting the power transmitted by the aerial A as P_1 and its power gain as G_1 and noting that in the VLW and LW bands use is ordinarily made of aerials having a low directivity in a horizontal

plane, Poynting's vector at point B (Fig. 5.3) will be given by

$$S = P_1 G_1 / A \text{ W/m}^2 \tag{5.11}$$

where A is the area over which the wave power is distributed. This is the side surface of a cone of revolution with axis OA and generatrix OC, bounded from above and below by spheres of radii a and $a + h$. Since $h \ll a$, the area A may be found by multiplying the circumference of the average radius CD by the altitude h as measured along the generatrix

$$A = 2\pi (CD)\, h = 2\pi (a + h/2)\, h \sin\theta \text{ m}^2 \tag{5.12}$$

The attenuation function F relative to free space may be found

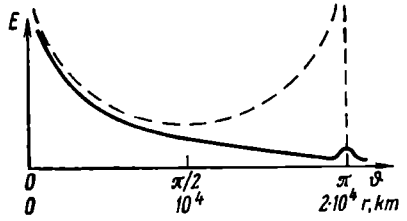


Fig. 5.4. Relationship between the wave field and distance without attenuation (the broken line) and with attenuation (the full line)

by equating Poynting's vector as given by (5.11) to $(P_1 G_1 / 4\pi r^2) F^2$

$$\frac{P_1 G_1}{2\pi (a + h/2) h \sin\theta} = F^2 P_1 G_1 / 4\pi r^2 \text{ W/m}^2 \tag{5.13}$$

whence

$$F = 2r / \sqrt{2 (a + h/2)\, h \sin\theta} \tag{5.14}$$

In the VLW and LW bands, it is convenient to describe the conditions of reception in terms of field strength. Using the already known expression for field strength in free space, extended to include the attenuation function, we get

$$E_{rms} = \frac{173 \sqrt{P_{1\text{kW}} G_1}}{r_{\text{km}}} F = \frac{245 \sqrt{P_{1\text{kW}} G_1}}{\sqrt{(a_{\text{km}} + h_{\text{km}}/2)\, h_{\text{km}} \sin\theta}} \text{ mV/m} \tag{5.15}$$

Eq. (5.15) is interesting in that it reveals an entirely unusual dependence of field strength on distance. This dependence is represented in Fig. 5.4 by the broken line whose shape indicates that the field strength decreases as the distance increases from zero to 10^4 kilometres (which corresponds to changes in the central angle from zero to $\pi/2$). At $r = 10^4$, the field is a minimum past which it again rises

in a symmetrical fashion so that at $r = 2 \times 10^4$ kilometres (that is, at the antipode) it is infinity. In the light of the assumptions made, this relationship appears natural, because the area over which the wave power is distributed is a maximum at $\theta = \pi/2$ and decreases with increase of distance. The rays travelling around the earth in all directions converge, as it were, at the focus of an optical system

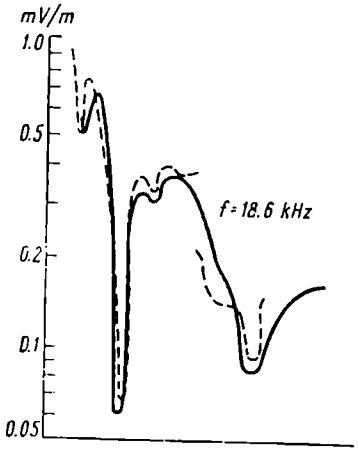


Fig. 5.5. Field as a function of distance according to calculation (the full line) and experiment (the broken line) at 18.6 kHz

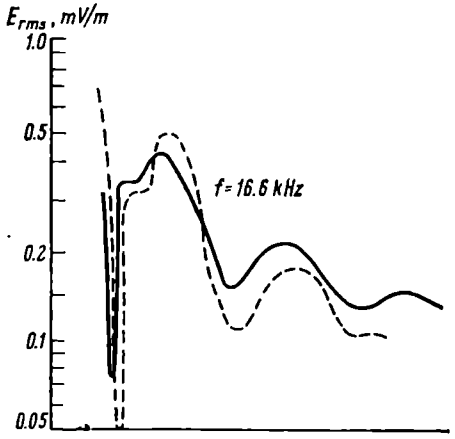


Fig. 5.6. Field as a function of distance according to calculation (the full line) and experiment (the broken line) at 16.6 kHz

at the antipode and, since we have neglected energy losses, the power flux density is the same as it was near the transmitting aerial.

Actually, however, the radio waves are markedly attenuated as they are reflected from both the ionosphere and the earth's surface. Besides, there is an asymmetry in the reflection conditions owing to the fact that on the daylight side of the globe the waves are reflected from the lower D region while on the night side they are reflected from the higher E region.

As adjusted for attenuation, the field-vs-distance curve takes the form shown in Fig. 5.4 by the full line. The curve shows no minimum at $r = 10^4$ kilometres any longer, but there is an increase in field strength at the antipode. This increase is observable under actual conditions and has come to be known as the *antipode effect*. It is caused by the multi-path interference of waves travelling around the globe in different directions.

The more rigorous theory of VLW and LW propagation, advanced by Alpert [82, 83] of the Soviet Union and by Wait [84] of the USA, is based on the normal mode method. The field at the point of recep

tion is treated as a sum of waves of different orders. In most practical cases, it is sufficient to select one mode of the first order. The equations thus derived have been used for computer calculations. The results check well with experimental data, as can be seen from the plots of Figs. 5.5 and 5.6 [84]. The full lines represent the relationship calculated by the normal-mode method, and the broken lines, the experimental data. Both the calculations and the experiments have been carried out for 18.6 kilohertz (Fig. 5.5) and 16.6 kilohertz (Fig. 5.6) for propagation over sea assuming $\omega_r = Ne^2/\varepsilon_0 m\nu$ to be 2×10^5, the ionospheric height $h = 70$ kilometres, and $\sigma \gg 1$ siemens per metre. It is to be noted that the calculated and experimental results check well for both short (under 3000 kilometres) and long (up to 6000 kilometres) distances. The effect of the earth's magnetic field has been tacitly neglected.

5.4. Variations in VLW and LW Propagation

On the whole, long waves are propagated under fairly constant conditions since they are reflected from the regular D and E regions. As a result, there are no sudden changes in signal strength, sudden radio black-outs, etc. Yet, there are more or less regular variations with time in LW propagation, which are examined in greater detail below.

Random Variations in Signal Strength. The electron density of the ionosphere fluctuates owing to irregularities in the ionizing stream and the ascending and descending motions of air masses in the upper atmosphere. As a result, the reflection height of the radio waves fluctuates, too. Because of this the relative phase relationship between ground and sky waves is not constant. The two waves drift in and out of phase. If the path difference is equal to any whole number of wavelengths, the two signals arrive at the receiver aerial in phase and reinforce each other. If the path difference is an odd number of half-wavelengths, the two waves reach the receiver in opposite phase and tend to cancel. The resultant signal strength can thus vary from a value given by the sum of the two to a value equal to their difference.

For $\lambda = 3000$ metres, the path difference of a half-wavelength is 1.5 kilometres, and for $\lambda = 20{,}000$ metres, it is 10 kilometres. It is obvious that such changes in the reflection height cannot be fast. This is why signal-strength fluctuations in the LW band are insignificant in depth (on the average, 10 to 30 per cent) and occur slowly (in a matter of tens of minutes or even hours), so that they cannot be heard in aural reception, although they can be detected in signal-strength records. As the wavelength is made shorter, these fluctuations grow more pronounced.

The Diurnal Variations. As a rule, the field strength rises at night, this increase being greater with shorter wavelengths. The underlying cause lies in the fact that the reflection from the *D* region entails larger energy losses than it does in the case of the *E* region.

The typical diurnal variations in field strength at wavelengths of 17,500, 11,650 and 5240 metres along a propagation path 5000 kilometres long are shown in the plot of Fig. 5.7. The distribution of

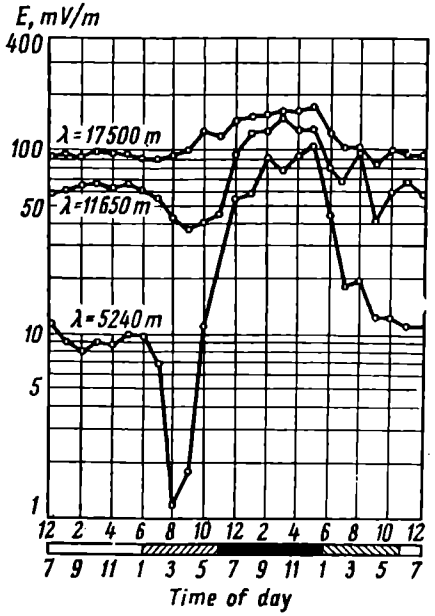


Fig. 5.7. Diurnal variations in field strength at wavelengths of 17,500, 11,650 and 5240 m along a propagation path 5000 km long

light and darkness along the propagation path is shown at the base of the plot. The blackened strip shows the hours when the entire propagation path is within the night zone.

Referring to Fig. 5.7, it is to be noted that at night the field strength increases almost ten-fold at $\lambda = 5240$ metres and only by a half at $\lambda = 17{,}500$ metres. Sometimes the field may undergo sudden changes at sunrise and sunset, as represented by the lower curve. The cause of these variations has not yet been identified. Apparently, these changes are due to the interference of the rays reflected at different heights—at the height of the *D* region on the day side and at the height of the *E* region on the night side.

The Seasonal Variations. In the LW band these variations are insignificant and manifest themselves only in an increase of about 20 to 50 per cent in day-time field strength in summer as compared with winter. On the other hand, there is a reduction in night-time field strength in summer as compared with that in winter.

The Sunspot Cycle Variations are likewise insignificant and take the form of an increase in day-time field strength with increasing solar activity. Cases have been reported of a twofold increase as solar activity changed from a minimum to a maximum.

The cause of this event apparently lies in the fact that the daytime increase in the electron density of the *D* region is accompanied by an increase in its conductivity and reflectance.

The Ionosphere Storm Variations. Ionosphere, or magnetic, storms mainly affect the upper reaches of the ionosphere which do not contribute to LW propagation. This is the reason why ionosphere disturbances affect but little LW propagation and are not accompanied by black-outs in radio communication.

There is a reduction in the received field during the main phase of magnetic storms and an increase during the recovery phase, that is, a few days later. This has led some investigators to think that magnetic storms can somewhat improve LW propagation. This is true, however, of ionosphere disturbances caused by solar flares, called 'SID'—*sudden ionosphere disturbances.* The improvement in LW propagation at the time of SID's is explained by the fact that long waves are reflected from the ionized layer formed at the height of the *D* region.

Utilization of Very Long Waves. The growing interest in VLW propagation in recent years has been stimulated by the fact that very long waves are propagated under extremely constant conditions—a fact which makes it possible to use very long waves (very low frequencies) in hyperbolic phase-comparison navigation systems. Both calculations and experiments show that the phase velocity undergoes changes (not over 3 per cent) within short distances from the transmitter and approaches that of light on distances in excess of 700 kilometres. The velocity of very long waves is discussed in detail in [83].

5.5. Calculation of Field Strength in the LW and VLW Bands

For engineering purposes, field strength in the LW and VLW bands is usually calculated by empirical formulae. The most commonly used one is the Austin-Cohen formula, originally derived on the basis of experimental investigations by Austin and Cohen for the U.S. Bureau of Standards.

The formula predicts with fair accuracy the day-time signal strength over sea for low-frequency radio ground waves (long waves in our classification) emitted by a vertical wire aerial. Since, however, the attenuation of long waves is mainly decided by energy absorption at the reflection from the ionosphere and is almost independent of the properties of the earth's surface, the Austin-Cohen formula may be used for LW propagation over land, provided that the propagation path is longer than 2000 to 3000 kilometres. In all cases, the Austin-Cohen formula may be used for distances not exceeding 16,000 to 18,000 kilometres.

When the transmitted power is known, the formula is

$$E_{rms} = \frac{300\sqrt{P_{1kW}G_1}}{r_{km}}\sqrt{\frac{\theta}{\sin\theta}}\,e^{-0.0014r_{km}/\lambda_{km}^{0.8}}\ \text{mV/m} \qquad (5.16a)$$

When the effective height h_e of the aerial and the current I_{rms} at its base are known, the formula is

$$E_{rms} = \frac{120\pi h_{e(m)}I_{rms\,(A)}}{\lambda_m r_{km}}\sqrt{\frac{\theta}{\sin\theta}}\,e^{-0.0014r_{km}/\lambda_{km}^{0.8}}\ \text{mV/m} \qquad (5.16b)$$

It is to be noted that, like Eq. (5.15) derived earlier, Eqs. (5.16a) and (5.16b) contain the term $\sqrt{\sin\theta}$.

From the Austin-Cohen formula it follows that for a given distance there is an optimum wavelength at which the field is a maximum. This can be proved by letting the effective height be independent of the wavelength and the current in the aerial be constant, differentiating Eq. (5.16b), and solving the equation $\partial E_{rms}/\partial\lambda = 0$. The solution will be the optimum wavelength. The physical cause for the existence of an optimum wavelength lies in the fact that as the wavelength increases the radiating ability of the aerial decreases (which leads to a reduction in the field strength) while the losses are minimized (which leads to an increase in the field strength). The optimum wavelength increases with increasing distance.

In practice, the radiation resistance of the aerial is frequency-dependent, and the current, instead of remaining constant, varies across the band. Besides, the conditions of reception are evaluated in terms of the signal-to-noise ratio and not in terms of field strength. Since the noise level varies with frequency, the optimum wavelength is that for which the signal-to-noise ratio is a maximum.

In order to determine the optimum wavelength, the power fed to the aerial is deemed constant across the band, the radiated power or the aerial current is found, and the field is determined for a number of wavelengths from 3000 to 20,000 metres, using Eqs. (5.16a) or (5.16b). On the other hand, knowing the dependence of the noise field on wavelength, the ratio E_s/E_n is found, and the wavelength for which this ratio is a maximum is chosen as the optimum one.

Example 5.2. Given: power fed to aerial, 400 kW; length of the radio circuit (or propagation path), 8000 km. To find: the optimum wavelength and the field strength due to a transmitter operating at that wavelength.

It is assumed that the dependence of the aerial efficiency (that is, the ratio of the radiated to the applied power) on frequency is known. Besides, it is important to know the dependence of the noise field at the point of reception on frequency for a specified hour of day and for a specified acceptance band of the receiver. These relationships are summarized in Table 5.1 along with the values of radiated power. Methods for determining the noise level are discussed in Chapter 6.

Solution: Substituting the radiated power from the table in (5.16a) gives the signal field (the upper line in Table 5.2). Now the signal-to-noise ratio is found (the second line of Table 5.2). From reference to Table 5.2, it is seen that the optimum frequency is 20 kHz, and the optimum wavelength is 15,000 m. At the same time, it is clear that the peak is not sharp, and the radio circuit may well use frequencies in the range from 20 to 40 kHz.

Operating at the optimum frequency, the transmitter will produce a received field of 36 μV/m.

Table 5.1. TO EXAMPLE 5.2

Frequency, kHz	15	20	30	40	60	80
Aerial efficiency, per cent	7.9	14	26	43	47	50
Daylight noise field, μV/m	40	30	25	18	10	5
Radiated power, kW	31.6	56	104	172	188	200

Table 5.2. TO EXAMPLE 5.2

Frequency, kHz	15	20	30	40	60	80
Signal strength, μV/m	38	36	26	20	8.4	3.6
Signal-to-noise ratio	0.95	1.20	1.15	1.13	0.84	0.72

In conclusion, it may be noted that the state of the troposphere has no marked effect on LW and VLW propagation. Nor are there tropospheric waves in these bands, because long-distance propagation within them is by way of the ionosphere. Long and very long

ground waves appear only within a short distance of the transmitter. At distances upwards of about 500 kilometres, the dominant mode is ionospheric waves.

5.6. Physical Processes in the Propagation of Medium Waves

In our classification, the term 'medium waves' applies to wavelengths from 1000 down to 100 metres (which corresponds to frequencies from 300 kilohertz to 3 megahertz).

While long radio waves are reflected from the lower edge of the ionized region without penetrating it for any considerable distance, the reflection of medium waves calls for a greater electron density.

Using Eqs. (4.104) and (4.103), describing the earthward reflection of rays at vertical and oblique incidence respectively, it is an easy matter to find the electron densities necessary for the reflection of medium waves.* These values for the E region are summarized in Table 5.3. The term 'oblique' applies to a ray touching the earth's surface. As is seen, its reflection calls for a minimum electron density.

Table 5.3

Frequency	Electron density, per cm³	
	vertical ray	oblique ray
300 kHz	10^3	40
3 MHz	1.2×10^5	3.6×10^3

From Table 5.3 it follows that in the high-frequency region of the band for a wave to be reflected it must reach a level where the electron density is several thousand per cubic centimetre. Because of this, we may conclude that medium waves trace out a curved path in the ionosphere, as shown in Fig. 5.8. In other words, medium waves are subject to gradual refraction in the lower sections of the trajectory and undergo total reflection at its crest in the ionosphere.

Medium waves suffer a marked attenuation in the ionosphere. Calculations show that the day-time attenuation in the D and E regions must be much greater than the night-time attenuation in the E region and must increase with decreasing wavelength.

* Calculations show that owing to a high collision frequency (about $5 \times \times 10^7$ per second) the electron density in the D region is insufficient to cause reflection of medium waves. Therefore, it may be assumed that medium waves are reflected solely from the E region. As before, the collision frequency in the E region is set at 10^6 per second.

Observations have fully confirmed the calculations. By day, wavelengths from 200 to 2000 metres are attenuated in reflection from the ionosphere to a point where ionospheric propagation may be fully ignored with existing transmitter powers. Conversely, the nighttime attenuation of these wavelengths is so small that ionospheric propagation is a dominating mode.

Wavelengths in the range from 2000 to 3000 metres, adjacent to the LW band, are not attenuated in reflection from the ionosphere by day so much as the shorter waves in the MW band are. In this respect, they stand midway between the wavelengths in the range 200-2000 metres and long waves.

A very important thing about medium waves is that they are propagated as ground waves by day and as both ground and ionospheric (sky) waves at night. It is only in winter and at high geographical latitudes that sky waves may produce a marked field at the point of reception also by day.

5.7. Variations in MW Propagation

Fading. Suppose the receiver (at *B* in Fig. 5.8) is within the range of the ground wave, *1*. Then at night it will pick up both the ground wave and the sky wave, *2*. Obviously, the sky wave and the ground wave travel by different paths of different distances to reach the same place. Under the circumstances, as has already been noted, the two waves interfere and produce a resultant signal strength which may vary from the sum to the difference of the two, according to the relative phase relationship between them.

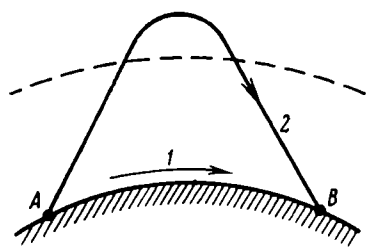


Fig. 5.8. Propagation of medium waves

If the ionosphere were time-invariant, the relative phase of the two waves would be constant and the resultant signal strength would have a time-invariant amplitude. Actually, the irregularities in the ionizing stream and air currents cause the electron density of the ionosphere to fluctuate. Because of fluctuation, the reflection height is anything but constant, and the path length for the sky wave is changing continually. As a result, the relative phase relationship between the two waves varies in a haphazard and unpredictable manner, causing a similar fluctuations in the amplitude of the resultant signal strength; this is amplitude fading.

A typical record of fading for $\lambda = 350$ metres at night appears in Fig. 5.9.

A change of Δr in path length for the sky wave produces a phase change of $\Delta\varphi = 2\pi\Delta r/\lambda$. It is readily seen that the same change of path length will produce a greater change of phase as the wavelength is made shorter. Furthermore, the shorter wavelengths are subject

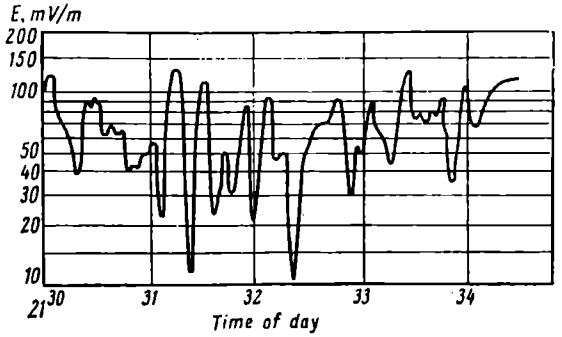


Fig. 5.9. Typical record of fading at $\lambda = 350$ m

to deeper fading but of a shorter duration. This is why fading is especially noticeable at wavelengths lying close to the lower limit of the MW band, that is, near 200 metres. The resultant signal strength may change tens of times, with the changes occurring inside, from a second to tens of seconds.

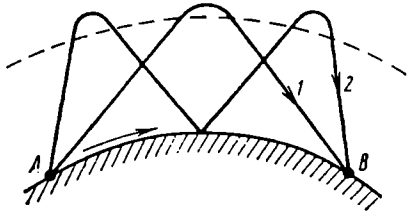


Fig. 5.10. Fading caused by the interference of sky waves experiencing one or two reflections

If the receiver is outside the range of the ground wave, the resultant signal is subject to fading caused by the interference of sky waves undergoing a different number of reflections from the ionosphere. Fig. 5.10 shows a case where the fading is caused by the interference of two waves one of which has been reflected from the ionosphere only once, and the other, twice. In other respects the mechanism of fading is the same as before.

Fading is detrimental to normal operation of radio circuits. This is why special measures are taken to minimize it. One is to provide

automatic gain control (*AGC*) in receivers. With *AGC*, a substantially constant output is automatically maintained despite a variation in the amplitude of the signal at the input. Another is to use so-called *anti-fading* aerials (this is widely practised by broadcast stations).

Consider operation of an anti-fading aerial in greater detail.

In an ordinary transmitting aerial, the radiating element is the vertical wire. The radiation pattern of such an aerial in a vertical plane is, to a first approximation, the shape of a semi-circle. In Fig. 5.11 it is shown by the full line, plotted on the assumption that

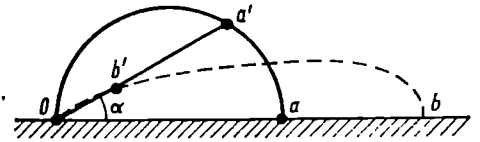


Fig. 5.11. Vertical radiation pattern of ordinary broadcast aerials (the full line) and communications aerials (the broken line)

the earth's surface is a sufficiently good conductor. In addition to waves propagated along the earth's surface, such aerials also emit waves at fairly large angle to the horizontal, that is, sky (or ionospheric) waves. If no measures were taken, the interference of the two waves, ground and sky, at the point of reception would be inevitable, resulting in heavy fading.

However, the aerial can be arranged so that the radiation pattern will make a small angle with the earth's surface, as shown in Fig. 5.11 by the broken line. Such an aerial will predominantly emit ground waves. From 5.11 it follows that the signals due to the ground waves emitted by the modified and the usual aerials are related as *Ob* to *Oa*. On the other hand, the signals due to the sky waves emitted by the two aerials are related as *Ob'* to *Oa'*. From the two ratios it is seen that the modified aerial produces a stronger ground-wave signal and minimizes that due to the sky waves.

Assume that the reception of broadcasts is free from fading when the ground-wave signal is at least three times as strong as the sky-wave signal. Assume further that for a conventional aerial this ratio of the two signal occurs at a distance r_1 from the transmitter. That is, a circle of radius r_1 will be a fading-free area for the conventional aerial. For the modified aerial, the fading-free area will then be of radius r_2. For practical aerials, the ratio r_2/r_1 is usually two or three.

The aerial can be modified to minimize the sky-wave signal by lifting the radiator higher above the earth's surface. With a well-conducting surface, use may be made of the image aerial, as shown

in Fig. 5.12. Then the rays due to the radiator and its image will add together along the ground, while in the direction making an angle α with the earth's surface there will be a phase difference $\varphi = (2\pi/\lambda)\, 2h \sin\alpha$ between the direct and the image rays owing to the path length difference $l = 2h \sin\alpha$. With a properly chosen height of the radiator above the earth's surface this phase difference will attenuate the sky-wave signal, and the radiation pattern of the aerial will make a small angle with the horizontal.

The Diurnal Variations in Signal Strength. This type of variation is extremely pronounced in the MW band and manifests itself in

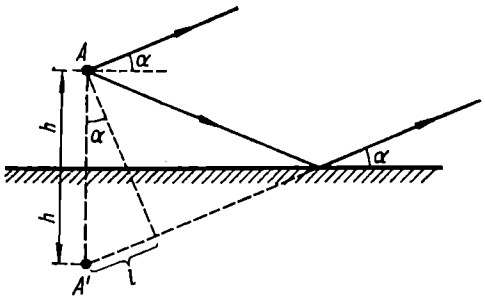


Fig. 5.12. An elevated radiator with a radiation pattern making a small angle with the horizontal

that the sky wave is completely or almost completely absorbed by day (as was already noted in § 5.6). At night, the sky waves are free to propagate. As a result, three cases of interference may present themselves, according to the distance from the transmitter.

If the receiver is close to the transmitter, the ground-wave signal will be stronger than the sky-wave signal even at night, and the signal strength will be practically independent of the time of day.

At greater distances from the transmitter, the signal strength will be decided by the ground waves by day, while at night the sky-wave signal will be stronger than the ground-wave one, and there will be fading.

At large distances from the transmitter, the ground-wave signal will be extremely weak even by day and there will be no reception. On the other hand, the sky waves may build up a strong signal at night. The situation is then one of no reception by day and good reception at night.

It should be added that during winter at high geographical latitudes the sky waves are not attenuated so strongly as they are at low latitudes and in summer. As a result the sky waves may well be propagated by day. Because of this, there will be fading in the first

two cases quoted above, and a substantial increase in signal strength in the third case.

The Seasonal Variations in Signal Strength. As is noted in § 4.6, the night-time electron density in the E region, which is the reflecting layer for medium waves, is almost independent of the season. Conversely, the daylight electron density of the E region rises in summer and falls markedly in winter.

Because of this, one may expect (and this has been proved experimentally) that there will be only a slight increase in the attenuation of the signal at night in summer as compared with the night-time in winter. Conversely, if the sky waves produce a strong signal by day (such as in the case of radio circuits at high latitudes), the sky-wave signal in the daytime will be much more attenuated in summer than in winter.

Speaking of the seasonal variations in signal strength, it is important to note that reception is evaluated in terms of the signal-to-noise ratio and not in absolute terms. Hence, the marked difference observed in the reception of medium waves in summer and winter is to be attributed not so much to variations in the signal field as to variations in the noise level (see Chapter 6). The point is that at the temperate and high latitudes of the Northern Hemisphere local thunderstorms occur only in summer, and so there is a substantial increase in noise level exactly in summer. In winter, the situation reverses, and the signal-to-noise ratio improves considerably.

The Sunspot Cycle Variations. The eleven-year cycle of solar activity has but an insignificant effect on the propagation of medium waves, consisting mainly in that the attenuation of the waves increases somewhat with increasing solar activity.

The Ionosphere Storm Variations. As with long waves, ionosphere storms have practically no effect on MW propagation.

The Effect of the Troposphere. Tropospheric refraction somewhat extends the range of the ground waves in the MW band and hardly affects the sky waves in that band. Besides, medium waves do not propagate in the tropospheric mode.

5.8. Calculation of the Sky-wave Signal in the MW Band

This section deals with the calculation of the MW signal received as sky waves reflected from the ionosphere, that is, of the signal produced at night-time in areas within the reach of the sky-waves.

Analysis of data on the strength of the signals from a number of broadcast stations, measured by the UER (European Radio Union) over a period of eight years, has yielded the following empirical

formula [85]

$$E_{rms} = \frac{10{,}233}{\sqrt{r_{km}}} \sqrt{P_{1kW} G_1}\, e^{-8.94 \times 10^{-4} \lambda_{km}^{-0.26} r_{km}} \ \mu V/m \qquad (5.17)$$

Equation (5.17) gives the annual median field strength for local midnight at the centre of the propagation path and a magnetic declination of 61° at that point. It is assumed that the relative sunspot number is zero. Use is made of a vertical aerial whose height is a small fraction of the wavelength.

In contrast to the opinion widely held in the past that the signal strength in the MW was independent of frequency, Eq. (5.17) points to this relationship, although a weak one and existing only at short distances from the transmitter.

Re-writing (5.17) as

$$E_{rms} = A/e^{a/\lambda} \ V/m \qquad (5.17a)$$

we can readily note that the shorter wavelengths (within the MW band) will produce a weaker field.

The UER publication [85] carries plots and charts with which the signal field, as given by Eq. (5.17), can be adjusted for the actual sunspot number, local time at the centre of the propagation path, magnetic declination at that point, and the radiation pattern of the aerial.

Example 5.3. Calculate the median signal field at night, at $r = 1500$ km from a transmitter operating at $\lambda = 900$ m and radiating 12 kW. The aerial is such that the radiation pattern in the vertical plane is nearly a semi-circle.

Solution: $0.9^{-0.26} = 0.97$

$$8.94 \times 10^{-4} \times 0.97 \times 1.5 \times 10^3 = 1.3$$

Now, using (5.17), $e^{1.3} = 3.65$ for a radiated power of 1 kW we get

$$E_{rms} = \frac{1.0233 \times 10^4}{\sqrt{1500 \times 3.65}} = 72 \ \mu V/m$$

For a radiated power of 12 kW,

$$E_{rms} = 72 \sqrt{12} = 450 \ \mu V/m$$

5.9. Physical Processes in the Propagation of Short Waves

The term 'short waves' applies to wavelengths from 100 down to 10 metres (which corresponds to frequencies between 3 and 30 megahertz). Like long and medium waves, short waves can be propagated in the ground and sky modes.

However, transmitters in the SW band produce little ground-wave signal because at these frequencies the ground wave is attenuated

rapidly and its range is short, just a few tens of kilometres. Therefore, SW transmitters rely more on the sky-wave signal which is capable of multiple reflections from the ionosphere and can cover distances however long because they experience an insignificant attenuation which decreases with increase of frequency.

Under normal conditions, the *D* and *E* regions absorb the energy of the sky wave in the SW band, and the *F* region reflects it back to earth, as shown in Fig. 5.13. From reference to the figure, we can readily conclude that under normal conditions the electron density of the *E* region is insufficient to reflect the short waves. On the other hand, the attenuation in reflection from the *F2* layer is considerably smaller than that experienced as the wave travels through the *D* and *E* regions both ways.

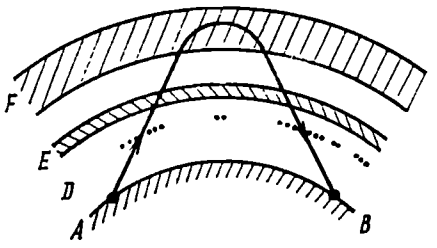


Fig. 5.13. Normal propagation of short waves

The former statement is obvious and is based on the well-known experimental finding that the daylight electron density of the *F2* layer is about ten times as great as it is in the *E* region. The latter statement needs a clarification.

As they travel through an imperfectly conducting medium, such as the ionized air, short waves experience attenuation due to the absorption of energy. The extent of attenuation is characterized by the exponential factor $e^{-\delta l}$ included in the expression for the signal strength, where δ is the absorption coefficient and l is the distance travelled by the wave.

Since the ionosphere is an inhomogeneous medium, δ will vary from point to point along the propagation path, and the total absorption along the entire propagation path will be given by the integral expression

$$e^{-\int_l \delta dl}$$

where the integral is to be taken over the entire path travelled by the waves in the ionosphere.

As is shown in § 1.5 the absorption coefficient δ is given by

$$\delta = p\omega/c \text{ m}^{-1} \tag{5.18}$$

where

$$p = \sqrt{\frac{1}{2}\left[-\varepsilon_l' + \sqrt{(\varepsilon_l')^2 + (60\lambda\sigma_l)^2}\right]} \tag{5.19}$$

In (5.19), ε_i' and σ_i are, respectively, the permittivity and conductivity of the ionized air (the effect of the earth's magnetic field being ignored), as defined by (4.54) and (4.55) which are written again for convenience

$$\varepsilon_i' = 1 - 3190N/(\omega^2 + \nu^2) \tag{5.20}$$

and

$$\sigma_i = 2.82 \times 10^{-8} N\nu/(\omega^2 + \nu^2) \text{ S/m} \tag{5.21}$$

where N = electron density, that is, the number of free electrons per unit volume (one cubic metre) of air;

ν = collision frequency (number of collisions between electrons and neutral molecules per second).

With the values that N and ν have in the E region and the $F2$ layer and in the SW band, the displacement-current density is much greater than the convection-current density. Writing this as $60\lambda\sigma_i \ll \varepsilon_i'$, we may considerably simplify Eq. (5.19).

After simple manipulations, we get

$$\delta = 60\pi\sigma_i/\sqrt{\varepsilon_i'} \text{ m}^{-1} \tag{5.22}$$

where

$$\varepsilon_i' \cong 1 - 80.8N/f^2 \text{ is the relative permittivity.} \tag{5.23}$$

In the SW band, the radian frequency is much greater than the collision frequency, because of which we may neglect ν^2 in comparison with ω^2 in (5.20).

Besides, as will be shown shortly, the value of ε_i' for oblique rays is very close to unity and is less than unity only at the point where the ray is turned to earth, that is, at the crest of the trajectory. Therefore, as a first approximation, we may set $\varepsilon_i' \cong 1$ in (5.22). Substituting (5.21) in (5.22) and noting that $\omega^2 \gg \nu^2$, we get

$$\delta \cong 1.35 \times 10^{-7} N\nu/f^2 \text{ m}^{-1} \tag{5.24}$$

Eq. (5.24) is interesting in at least two respects. Firstly, it shows that, in contrast to the relation between the absorption coefficient and frequency for imperfect conductors, the absorption coefficient in the ionosphere is inversely proportional to the square of the frequency. Secondly, for a given frequency, the extent of absorption is decided by the product of the electron density and the collision frequency.

Ionospheric measurements have shown that at noon the maximum electron density in the $F2$ layer is 10^{12} per cubic metre, and for the E region, 10^{11} per cubic metre. The collision frequency is about 10^8 per second for the $F2$ layer and about 10^6 per second for the E region. The $N\nu$ product for the E region is thus 10^{17} against 10^{15} for the

F2 layer. In other words, the absorption coefficient for short waves in the *E* region is about a hundred times as great as it is in the *F2* layer or, which is about the same, the absorption in the *F2* layer may be neglected in comparison with that in the *E* region.

Under actual conditions, short waves are attenuated not only in the area of the peak electron density of the *E* region, but also in the lower reaches of the ionosphere, namely at the lower edge of the *E* region and in the *D* region. At these altitudes, the collision frequency is comparable with the radian frequency, and ν in (5.21) and (5.24) cannot be neglected in comparison with ω any longer. As a consequence, the absorption coefficient no longer varies in inverse proportion to the square of the frequency, although it still decreases with increase of frequency.

For radio communication using short waves two conditions must be satisfied simultaneously:

(1) the frequency used must be lower than the maximum usable value determined for a propagation path and for a given ionization of the reflecting layer;

(2) the absorption of waves in the *D* and *E* regions must not be too heavy. That is, the signal strength produced by ordinary transmitters and ordinary aerials should be sufficient for reliable reception.

The first condition limits the usable frequency band from above, while the second condition does so from below.

It is important to note that the first condition is a critical one because otherwise radio waves will not be reflected from the *F2* layer altogether and, despite any increase in transmitter power, will fail to reach the desired point of reception. The second condition is not so critical, because the use of a lower frequency may be compensated by an increase in transmitter power. Indeed, it is possible to maintain radio traffic on relatively low frequencies of the SW band by the use of high-power transmitters.

In view of the two conditions, it is customary to operate on wavelengths from 10 to 25 metres for long-distance traffic by day, 35 to 100 metres at night, and 25 to 35 metres in twilight hours. Of course, this division is an arbitrary one, as it is subject to adjustments for the time of season, phase of the eleven-year solar cycle, and geographical location.

As is seen, the "day-time" wavelengths are much shorter than the "night-time" ones. The point is that the day-time electron density of the *F2* layer is very high, and it can reflect the higher frequencies. On the other hand, these higher frequencies will experience an insignificant absorption in the *D* and *E* regions highly ionized by day. Conversely, the night-time electron density of the *F2* layer goes down, and the usable frequency must be also brought down in order to meet the first requirement. At night, however, the electron density

in the *E* region is reduced while the *D* region disappears altogether, because of which the attenuation will be insignificant even for the lower frequencies of the SW band.

Figs. 5.14 and 5.15 show what happens when any one of the two conditions is not satisfied. If a station keeps on using one of its "day-time" frequencies at night, the instant may arrive when the

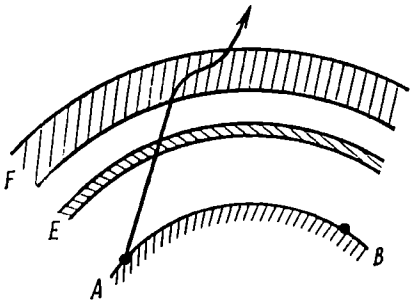


Fig. 5.14. Propagation with the transmitter frequency exceeding the critical frequency

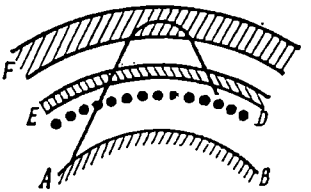


Fig. 5.15. Propagation under conditions of increased absorption in the *D* and *E* regions

electron density of the *F2* layer drops to a value at which the transmitter frequency exceeds the critical frequency, and the wave ceases to be reflected from the ionized layer. As is shown in Fig. 5.14, the ray passes through the *F2* layer.

If a station keeps on using one of its "night-time" frequencies by day, this frequency is obviously below the critical value and it will be reflected from the *F2* layer. However, in passing through the *D* and *E* regions, it will experience a very strong attenuation, since, according to (5.24), the absorption coefficient drastically increases with decrease of frequency. The attenuation of "night-time" frequencies by day is usually so strong that the signal strength is insufficient for reliable reception.

The foregoing applies to normal conditions of SW propagation, that is, when the two basic regions of the ionosphere participate in it. The situation is different, however, during the periods of high solar activity in summer. At that time the ionization of the *E* region rises to a point where it often acts as a reflecting layer for radio waves about noon, while the *D* region serves as an absorbing layer.

Normal conditions for SW propagation may also be upset by the Sporadic *E* (*Es*) layer. The *Es* layer may occur at any time of day and may sometimes have a fairly high electron density. Since it is located at the height of the normal *E* region, that is below the *F2* layer, the *Es* acts as a reflecting layer for short waves.

Short waves offer a number of major advantages over long and medium waves, such as insignificant attenuation in the ionosphere. On the other hand, short-wave signals rely for refraction and reflection on the *F2* layer, a rather irregular structure, because of which they are not propagated under as steady conditions as medium, long and very long waves are.

The irregularities of the *F2* layer have a two-prong effect on SW propagation. Firstly, SW signals are subject to fading which is much more pronounced than it is in the MW band. Secondly, the *F2* layer is subject to marked diurnal and magnetic-storm variations, and so is the short-wave signal. There are also some other factors affecting SW propagation, such as the skip zone, echoes, and diffuse reflection.

5.10. Fading in the SW Band

Fading on short waves manifests itself as random (sometimes more or less regular) variations in signal strength between maxima and minima, as shown in Figs. 5.16 and 5.17.

Under fading conditions, the amplitude of the signal may vary by a factor of tens or oven hundreds. The period of fading, defined as

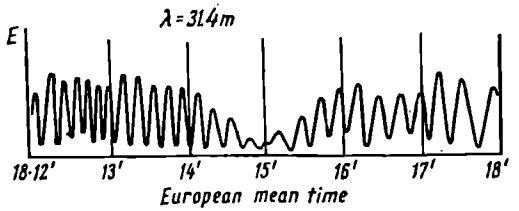


Fig. 5.16. Signal fading at $\lambda = 31.4$ m (after Shchukin)

the time interval between two consecutive maxima or minima, may be from tens of seconds to a few tenths of a second [10].

Fading may take a multitude of forms. In continuous-wave (CW) work, for example, the signal strength may vary more or less gradually. In other cases, fast changes in the signal are superimposed on the slower changes in the wave field.

As in the MW band, the main cause of fading is the interference of different wave components that travel slightly different paths in arriving at the receiver (*interference fading*). In contrast to the MW band, however, where fading occurs due to interference between ground and sky waves, in the SW band the interfering components are two or more sky waves from the same transmitter that have followed different paths (*multipathing*). This may be due to several causes.

Fig. 5.18 shows a case where of the two rays reaching the receiver one has been reflected once and the other twice between the ionosphere and earth. Since ionospheric height at the reflection points varies continually, the total path lengths also change, and the phase

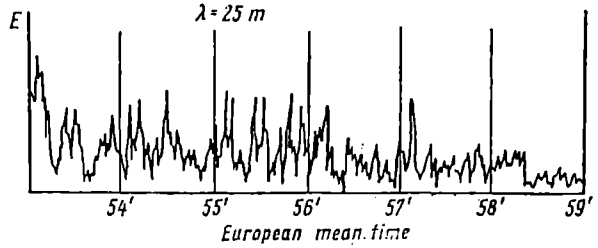


Fig. 5.17. Signal fading at $\lambda = 25$ m (after Shchukin)

relationship between the rays arriving at the receiver is varied continually and in a random manner. As will be recalled, if the path difference is an odd number of half-wavelengths, the two signals arrive at the receiver aerial in opposite phase and tend to cancel. For the path lengths involved, a half-wavelength works out to a few tens of metres. Obviously, fluctuations in the state of the ionosphere can produce such insignificant variations in the reflection height at

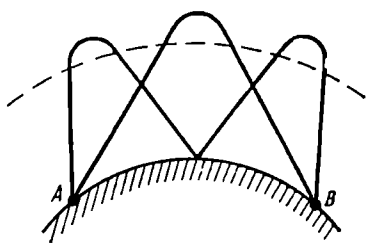


Fig. 5.18. Fading caused by the interference of rays experiencing different number of reflections

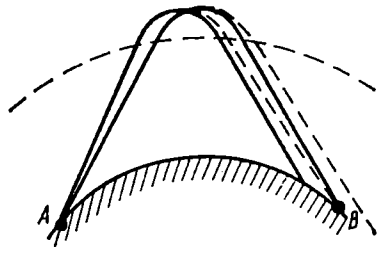


Fig. 5.19. Fading caused by the interference of the ordinary and extraordinary rays

any time. As a result, the signal strength will be a minimum. If the path difference happens to be equal to any whole number of wavelengths, the two rays will arrive at the receiver aerial in phase, and the signal strength may double.

Another cause for interference fading may be the fact that the earth's field splits up an incident short wave into two components—the ordinary wave and the extraordinary wave (Fig. 5.19), as is explained in § 4.9. Naturally, each strikes a different point on the earth's surface than the other. It is an easy matter to show that at

each point there will be both an ordinary and an extraordinary wave, but from different incident waves reflected from different areas in the ionosphere. In the general case, the two waves at each point will be in opposite phase, and again fading will result. It may be added that rays after a single and double reflection may interfere with the split-up components, in which case four waves will interfere with one another.

A more important case of interference fading is schematically shown in Fig. 5.20. The point is that the ionosphere is anything but

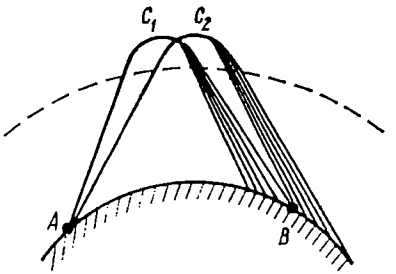


Fig. 5.20. Fading caused by the interference of elementary rays

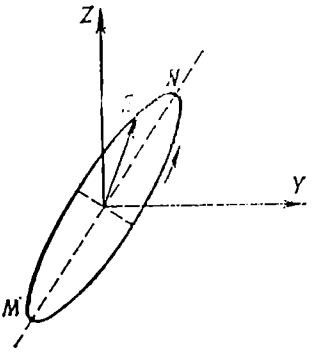


Fig. 5.21. Faraday fading

a smooth reflecting surface. Owing to the inevitable irregularities, the incident waves experience partly diffuse instead of specular reflections. As a result, the ray emerging from the ionosphere appears as a bundle of elementary rays.

Experiments have shown that the angular spread of such a bundle of rays may be from one to five degrees. As a result, a multitude of elementary rays belonging to different bundles may strike the same point on the earth's surface. Referring to Fig. 5.20, the rightmost elementary ray of the bundle C_1 interferes with the left-most ray of the bundle C_2. Since each ray continually changes in phase owing to fluctuations in the reflecting layer, the result is random variations in the signal strength, that is, fading.

In addition to interference fading, there is *Faraday fading*. It is a rare occurrence (about 10 to 15 per cent of all cases) and consists in the following.

As has been noted earlier, a plane-polarized wave entering the ionosphere is split up by the earth's magnetic field into two elliptically polarized components. Given a favourable electron density distribution in the ionosphere and orientation of the earth's magne-

tic field relative to the direction of wave propagation, the resultant field will be an elliptically polarized one, with a very oblong ellipse of polarization. Because of random fluctuations in the electron density along the propagation path, there will be continuous changes in the direction of the major axis MN of the ellipse of polarization (Fig. 5.21).

Now we assume that the signal is received with a vertical aerial. Obviously, the signal strength will be a maximum when the major axis of the ellipse is in the direction of the vertical. Conversely,

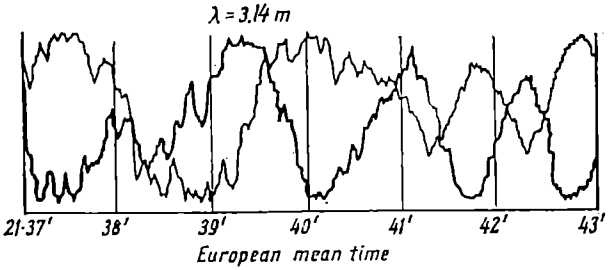


Fig. 5.22. Faraday fading. The heavy line represents the vertical electric-field component, and the fine line, the horizontal one (after Shchukin)

when the major axis MN of the ellipse is horizontal, the signal strength will be a minimum. The result will obviously be the fading of the total signal.

In the Soviet Union this form of fading has been investigated in detail by Shchukin [10] who simultaneously measured the vertical and horizontal components of the received field, using a vertical and a horizontal dipole each connected to a separate receiver by a transmission line. The voltages appearing across the receiver outputs were recorded photographically on a common film strip. An example of the record appears in Fig. 5.22. The heavy line represents the signal strength due to the vertical aerial, and the fine line, to the horizontal one. As is seen, an increase in signal strength due to the vertical aerial is always accompanied by a decrease in that due to the horizontal aerial, and vice versa, which is an indication that there are continuous changes in the orientation of the major axis of the polarization ellipse.

Since Faraday fading also involves interference between two waves, it may be rightly treated as a special case of interference fading by use of the same quantitative methods.

Deep fading in the SW band distorts the output of the receiver badly. This is why suitable measures have to be taken in order to control it.

As in tropospheric propagation, fading proper ought not to be confused with slow random variations in the average field strength. While fading proper is described by the Rayleigh distribution, the slow variations have the logarithmically normal distribution. Measurements have shown that with the logarithmically normal distribution the standard deviation is 8 decibels [33]. The density function and distribution function for the Rayleigh fading are given by (3.68) and (3.69), respectively. Everything said in § 3.11 about fading control in tropospheric propagation fully applies to the

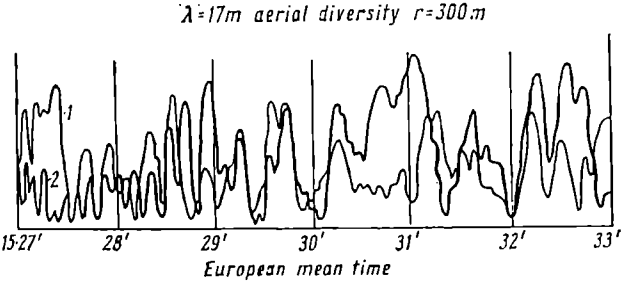


Fig. 5.23. Control of fading in space-diversity reception (after Shchukin)

SW band. However, because of the heavier radio traffic in the SW band, selective fading is hardly useful. The main method for fading control is diversity reception. The gain thus obtained is given by (3.77) and the plot of Fig. 3.28.

According to [53], the correlation time τ, as found from the auto-correlation characteristic, Eq. (3.74), during which the auto-correlation coefficient drops to $\rho = 0.5$ to 0.7 lies in the interval from 4.5 to 1.5 seconds. When the receiving aerials are spaced 10 wavelengths apart (in the direction of propagation), the cross-correlation coefficient decreases to $\rho = 0.4$. The effect of space-diversity reception, as this method is called, is shown in the plot of Fig. 5.23.

5.11. The Primary Skip Zone

The term *primary skip zone* (also known as a zone of silence) applies to the annular belt around a radio transmitter where its signals cannot be received. It extends beyond the ground-wave range (*AB* in Fig. 5.24) but lies within the so-called *skip distance*, or the distance from the transmitter to a point where a specified radiated frequency first returns to earth as a sky wave (*AC* in Fig. 5.24).

In this zone, the ground-wave signals are attenuated below a usable level. As to the sky waves, the radiation with an angle of elevation, β, exceeding a certain critical value (dependent on the wave frequency) penetrates the ionosphere. Thus, the waves at and near the vertical are not reflected back to earth. On the other hand, the radiation with an angle of elevation equal to this critical value is turned back to earth at point C, and the area surrounding the trans-

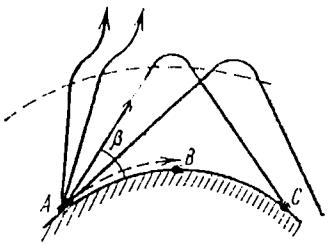


Fig. 5.24. Skip zone and skip distance

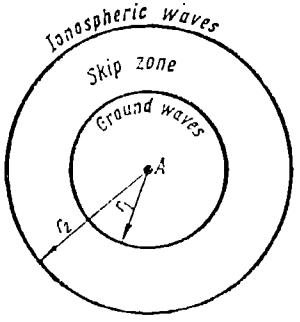


Fig. 5.25. Skip zone in the propagation of short waves

mitter at that distance receives some returned sky wave. Within the distance BC, there are, thus, neither ground-wave nor sky-wave signals.

Fig. 5.25 shows the primary skip zone in plan on the assumption that the transmitting aerial is radiating equally well in all directions and at all angles to the horizontal and the propagation of sky waves is independent of azimuth. Otherwise, the primary skip zone (or simply the *skip zone*) will be a more complicated annulus.

The inside radius of the skip zone is decided by the ground-wave range. For a given transmitter power, it is independent of the time of day and depends solely on frequency — the higher the frequency radiated, the smaller the inside radius of the skip zone. The absolute value of the inside radius can be found by calculating the strength of the ground-wave signal, using the methods set forth in Chapter 2.

The outside diameter is decided by the skip distance, that is, the conditions of ionospheric propagation. Because of this, it will vary with the time of day, season, and frequency. For a given frequency, the night-time electron density decreases, and for f_{max} given by

$$f_{max} = \sqrt{\frac{80.8 N_{max}(1 + 2h/a)}{\sin^2 \beta + 2h/a}} \text{ Hz} \qquad (5.25)$$

to remain unaffected, the angle β must be decreased. As a consequence, more oblique waves, that is, those with a smaller angle of ele-

vation, can be reflected from the ionosphere at night. Of course, they are returned to earth at greater distances from the transmitter and the skip distance increases.

The same equation, (5.25), answers why the outside diameter of the skip zone, or the skip distance, varies with frequency (for a given maximum electron density). If f_{max} increases while N_{max} is held constant, the equality will be maintained if the value of β is decreased. But this is another way of saying that the skip distance, or the outside diameter of the skip zone, increases with frequency.

To sum up, as the frequency radiated by the transmitter is raised, the skip zone increases in area due to both the accompanying decrease in the internal radius (the ground-wave range) and the increase in the outside radius (the skip distance). As we move towards the medium waves, the ground-wave range and the skip distance become equal at first and then overlap each other, so that there is no skip zone left. This explains why there is no zone of silence in transmission on medium waves.

Methods for determining the skip distance are discussed in the section on the maximum usable frequency.

5.12. Short-wave Echoes

Because of the relatively low attenuation, short-wave signals may cover exceptionally long distances. Indeed, they may travel all round the earth by a series of multiple reflections between the earth and the ionosphere. This is the so-called round-the-world echo.

There may be forward and backward round-the-world echoes. Fig. 5.26 shows a transmitter, *A*, and a receiver, *B*. Normally radio waves are propagated along the shortest path *AB* in the forward direction *1* (we have omitted the multiple reflections so as not to overcrowd the diagram). Given favourable conditions (to be discussed shortly), the receiving aerial at *B* may pick up the transmitted ray that has travelled all the way round the globe in the backward direction *2*. If the shortest distance between transmitter and receiver is less than 20,000 kilometres, the signal travelling in the opposite direction will arrive at the receiver some time after the main signal. This delay increases with increase of difference in length between the forward and backward propagation paths.

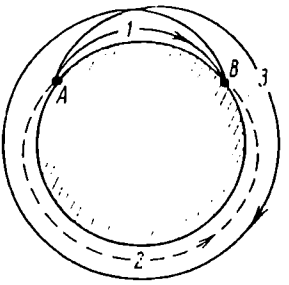


Fig. 5.26. Round-the-world echo

It takes radio waves about 0.13 second to circle the earth at the equator with the velocity of light in free space, and so the delay between the forward and backward signals is, to a first approximation, equal to $0.13 \div 40 = 0.003$ second for each thousand kilometres of the path length difference.

Sometimes the forward signal may arrive at the receiver after it has travelled round the globe in the forward direction (ray *3*). This will be a forward round-the-world echo. Its delay will be about 0.13 second.

Both forward and backward echoes may be multiple, in which case the delay between the two signals will be a multiple of 0.13 second. Because of this delay, echoes can be discerned in the aural reception of both telegraph and voice signals.

In some cases an echo may be comparable in strength with the wanted signal, and its detrimental effect will then be only too obvious. In telegraph working, it may give rise to spurious signals, while in voice operation it will present itself as sustained reverberation or an acoustical echo. In fascimile, the round-the-world echo may do a good deal of harm, too. This is why every effort should be made to control it.

The backward round-the-world echo is simplest to suppress. For example, this can be done by the use of isotropic aerials for both transmission and reception. The forward round-the-world echo is much more difficult to control. Obviously, isotropic aerials would be of little help here, because both the wanted signal and the echo travel in the same direction and in many cases at nearly the same angles to the horizontal.

Control of the forward round-the-world echo may be based on the fact that despite the same illumination at reflection points at the instants when a round-the-world echo is produced the state of ionization is different there because of the difference in the duration of day and night at those points. An obvious way out is to select radio frequencies which will suffer a greater attenuation or will not be reflected by the ionosphere at all when propagated along the longer path. However, it is not at all unlikely that the frequency will have to be changed for two or three hours while the echo exists.

In cases where the wanted signal considerably exceeds the echo in strength, the desired effect can be secured by the use of a lower frequency which is subject to greater attenuation.

Experiments [86] have shown that the time of travel around the globe for the signal is subject to marked variations. In the 218 measurements taken, the time of travel round the earth varied from 0.13760 to 0.13805 second.

In addition to the round-the-world echo, short waves are subject to the so-called near or direct echo. It occurs when two or more rays

sent out by a pulsed transmitter arrive at the receiver after a different number of reflections from the ionosphere. From reference to Fig. 5.27, it is readily seen that the near or direct echo is produced by about the same mechanism as interference fading. With fading,

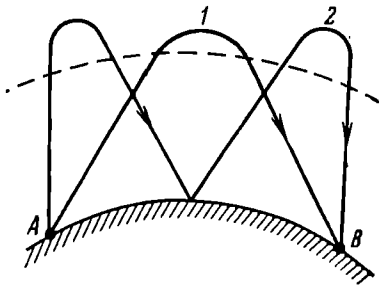


Fig. 5.27. Near echo

however, the signal of long duration (from 2 milliseconds upwards) travels along path *1* and arrives at the receiving aerial simultaneously with the signal that has travelled along path *2*. The signals overlap, as shown in Fig. 5.28*a*, and fading takes place. With

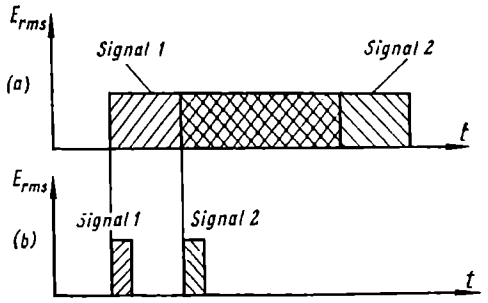


Fig. 5.28. Explaining the origin of near echo

a direct echo, the duration of the signal is not more than one millisecond, the two signals do not overlap, and there is no fading (Fig. 5.28*b*).

Simple calculations would show that the time delay of the second signal relative to the first ranges between 0.5 and 2 milliseconds. The near echo interferes with operation of fascimile and high-speed telegraphy circuits.

The near echo may be controlled by use of (1) frequencies so close to the critical one that the rays nearest to the vertical will not be

reflected; and (2) highly directional transmitting and receiving aerials with radiation patterns at low angles to the horizontal to minimize sky-wave signal strength.

5.13. The Sunspot-cycle Variations in Short-wave Signal Strength

Astronomers have observed that solar activity varies periodically. Variations in solar activity manifest themselves in the varying number and size of sunspots, faculae (sometimes called torches),

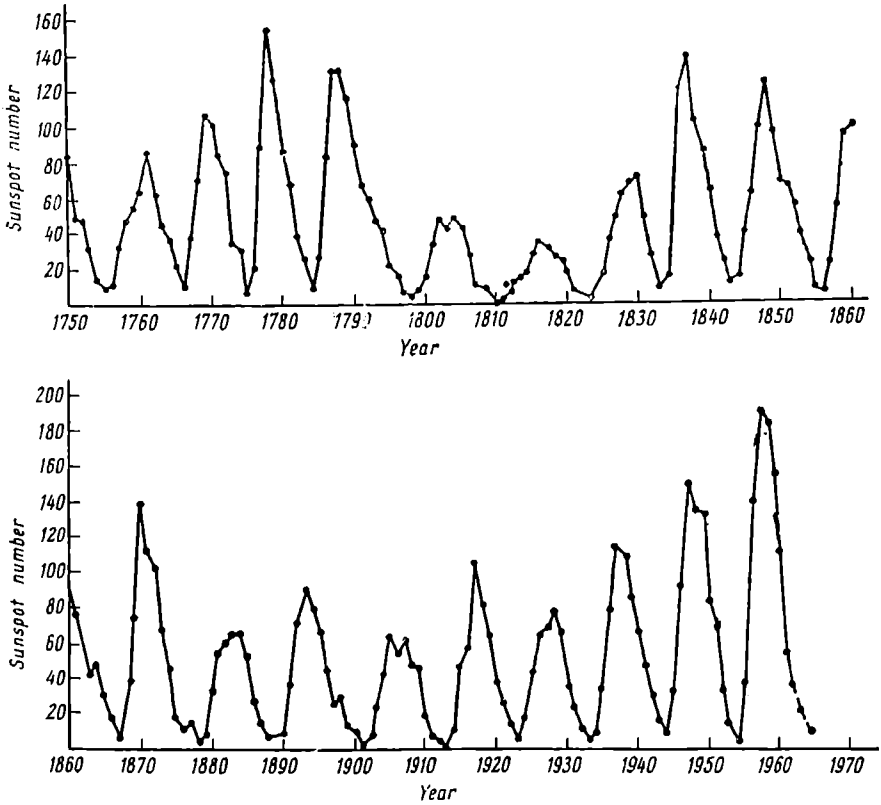


Fig. 5.29. Variations in the relative sunspot number between 1750 and 1964. The broken line represents the expected number

prominences, and flocculi, and changes in the intensity of solar radio emissions. It is customary to evaluate solar activity in terms of the sunspot (Wolf) number.

The manner in which the sunspot number varies can be visualized from reference to Fig. 5.29 which shows changes in the sunspot number between the years 1750 and 1964. During the intervening period, the minimum sunspot number was 0.0 (in 1810), and the maximum number was 190.2 (in 1957). The shape of the curve bears out that solar activity is not a strictly repetitive process. Yet, in the two hundred years covered by the plot, solar activity had an average period of about 11.3 years with a deviation of plus or minus four years from this average.

Of special interest are variations in solar activity in recent years and the state of the sun at present. Table 5.4 shows the maximum and minimum values of sunspot number since 1933, and Table 5.5 presents the sunspot numbers for the last ten years.

Table 5.4. MAXIMUM AND MINIMUM VALUES OF SUNSPOT NUMBER SINCE 1933

Year	Maxima	Minima	Cycles between maxima, minima, and maximum-minimum
1933		5.7	4, 11
1937	114.4		7, 10
1944		9.6	3, 10
1947	151.6		7, 10
1954		4.4	3
1957	190.2		

Table 5.5. SUNSPOT NUMBER SINCE 1953

Year	Sunspot number	Year	Sunspot number
1953	13.9	1959	159.0
1954	4.4	1960	112.3
1955	38.0	1961	53.9
1956	141.7	1962	37.5
1957	190.2	1963	27.9
1958	184.8	1964	10.5

The data of Table 5.4 show that the sunspot number increases at a high rate from a minimum to a maximum (during three or four years) and decreases at a relatively low rate from a maximum to a minimum (during seven years).

Sunspot field measurements have shown that their polarity reverses every eleven years. This is why astronomers think it more appropriate to speak of a 22-year cycle of solar activity.

Variations in solar activity are accompanied by changes in the strength of the short ultra-violet radiation (shorter than 2200 Ångstrom units), X-ray, radio emission, and corpuscular streams, that

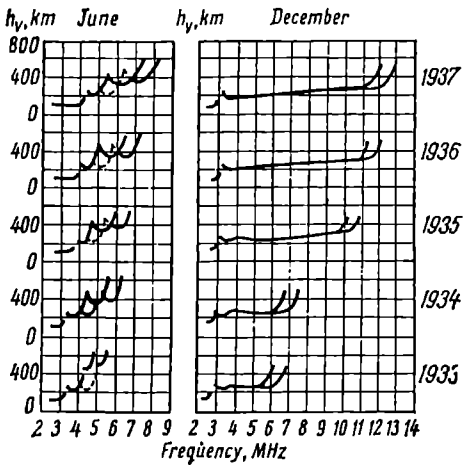


Fig. 5.30. The effect of the 11-year sunspot cycle on summer and winter average monthly ionograms

is, in the types of radiation that are completely absorbed by the upper atmosphere and have no chance of reaching the earth's surface. Because of this, measurements of solar flux density at the earth's surface fail to reveal the eleven-year cycle.

An increase in solar activity raises the electron density in all layers of the ionosphere. This rise is most pronounced in the outer reaches of the atmosphere, that is, the *F2* layer, and is least noticeable in the *E* region. Reliable data on the *D* region are lacking.

Fig. 5.30 shows June and December ionograms (plotted at I, at 40° N) for the period between 1933 and 1937 when the sunspot number varied between 5.9 and 114.4. It may be seen that the critical frequency for the *E* region had increased by 25 per cent during the period. During the same time the critical frequency for the *F2* layer had doubled. About the same relationship holds for the MUF. Since the higher frequencies experience a lower attenuation in the

D and *E* regions, the advisability of using higher frequencies is obvious. This is another way of saying that during the incidence of increased solar activity the "night-time", "twilight" and "day-time" frequency bands should be shifted in the direction of higher frequencies. There is also an overall improvement in short-wave propagation owing to a greater ionization of the upper atmosphere.

It has been noted that ionospheric prediction maps for the *F2* layer do take into account the effect of the eleven-year cycle.

A more detailed discussion of the effects that the solar cycles have on short-wave propagation may be found in [87] and [88].

5.14. The Effects of Geomagnetic Disturbances on Short-wave Propagation

Among the various forms of geomagnetic disturbances short-wave propagation is affected most of all by (1) ionosphere disturbances related to global magnetic storms, (2) absorption in auroral regions, (3) polar-cap absorption events (PCA), and (4) sudden ionosphere disturbances (SID).

Ionosphere Disturbances. These are produced by the periodic eruptions of charged particles from the sun. On breaking into the earth's atmosphere, they heat its upper layer and upset the normal ionospheric structure, mainly that of the *F2* layer. As they move down to the earth's surface and interact with the terrestrial magnetic field, the corpuscular streams deviate from the original straight-line paths to form eddies which concentrate in the polar regions. This is why ionosphere disturbances mostly occur in the annular zone around the magnetic poles. The intensity of ionosphere disturbances markedly falls off on moving into temperate geographical latitudes.

In addition to ionosphere disturbances, the influxes of solar particles bring about magnetic storms and auroral displays. In contrast to local storms observed solely in the polar regions, these global storms are registered all over the world. According to the latest findings, global magnetic storms are related to the radiation belts around the earth. It is very likely that the particles "dumped" into the earth's atmosphere come from the store of previously trapped particles in the radiation belts. Some investigators tend to attribute global magnetic storms to MHD waves. Global magnetic storms and the related ionosphere disturbances may last from several to forty-eight hours.

During the incidence of ionosphere disturbances the electron density of the *F2* layer takes a dip and the virtual heights go up (due to the heating), so that the maximum usable frequencies are decreased (in full agreement with Eq. 4.102). The regular structure of

the *F2* layer is disrupted, and strata appear in it. During very strong ionosphere disturbances the ionization of the *F2* layer may drop to a point where the layer will not reflect short waves any longer. In a sense, the *F2* layer is destroyed during ionosphere disturbances at high geomagnetic latitudes.

As solar activity increases, the intensity and rate of magnetic storms and ionosphere disturbances go up, as shown in Fig. 5.31 where years are laid off as abscissa, the average sunspot number R as ordinate on one side and the magnetic activity index u_1 as ordinate on the other side of the Y-axis. It is immediately seen that the two curves are nearly identical.

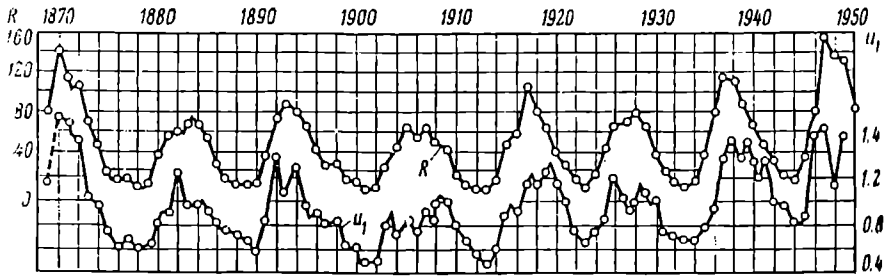


Fig. 5.31. Correlation between variations in the sunspot number and magnetic activity from 1870 to 1950

Strong ionosphere disturbances and magnetic storms follow a cycle of twenty-seven days (which is the sun's period about its axis). This is because sources of corpuscular radiation can remain active for months and will hold the same position relative to the earth after a cycle of 27 days, so that a disturbance or a storm will recur.

Now that there is a well-organised solar service, ionosphere disturbances and magnetic storms can be predicted with sufficient accuracy.

Disruptions of radio communication on short waves due to ionosphere disturbances can be avoided by use of (1) high-power transmitters and better transmitting and receiving aerials in regions subject to ionosphere disturbances (above geomagnetic latitude 50° North); (2) progressively decreasing frequencies after the onset of an ionosphere disturbance to counteract the decrease in the electron density of the *F2* layer (this method has been suggested by the Radio Moscow staff); and (3) short-wave relay stations located outside the disturbance area, as shown in Fig. 5.32. The last-mentioned method has been developed by a team of scientists under Shuleikin [89]. Referring to Fig. 5.32, AB is the main radio circuit lying within the disturbance area. A relay station is provided at C, south of the disturbance. The radio circuits AC and CB lie outside

the disturbance area, and so radio traffic can be maintained without interruption during a storm.

Radio-wave Absorption in the Auroral Zone. In addition to global magnetic storms, the polar regions are often the scene of local storms. These are accompanied by marked increases in radio-wave absorption, especially in the auroral zone—the annular or helical zone up to ten degrees wide at geomagnetic latitudes 67.5° N and S. In that zone, charged particles with energies up to one megaelectron-volt, penetrating the atmosphere as far as the *D* or *E* region, produce a strongly ionized region which absorbs short waves. On the other hand, the ionization of this layer is insufficient to reflect short waves. Radio-wave absorption in the auroral zone is often a sign of an advancing global magnetic storm. Auroral-zone absorption may last from hours to days.

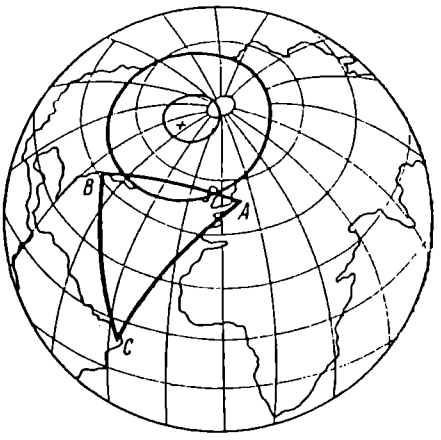


Fig. 5.32. Use of radio relay links to combat ionosphere disturbance

Polar Cap Absorption Events (PCA). In contrast to the auroral zone which is an annular region, a polar cap is a circular area with a geomagnetic pole (North or South, as the case may be) as centre, extending out to geomagnetic latitudes 64°.

It is believed that the recurring strong ionization in the polar caps is produced by the influx of solar particles with energies from 10 to 100 megaelectron-volts, that is, more energetic than those responsible for radio-wave absorption in the auroral zone. As in the latter, the charged particles produce a strongly ionized layer at the height of the *D* region throughout the polar cap. This layer absorbs short waves and may persist for ten to 300 hours.

Sudden Ionosphere Disturbances (SID). Direct measurements with the instruments carried by geophysical rockets have shown that solar flares give rise to a sudden increase in the intensity of the short ultra-violet (Lyman-alpha radiation, 1215.7 Angstrom units) and X-rays (1-1.5 Ångstrom units). Penetrating the earth's atmosphere very deeply, these radiations produce a strong ionization in the *D* region and, as a consequence, considerable absorption of short waves.

Experiments have shown that this increase in the ionization of the *D* region interferes with normal reflections from the *F2* layer

but it is insufficient to make the *D* region reflecting for short waves.

Propagated along straight-line paths and undeflected by the terrestrial magnetic field, ionizing radiant energy can reach only the daylight side of the globe, and so radio-wave absorption increases as the zenith angle of the sun decreases.

SID events are in no way related to global magnetic storms and usually subside in a matter of several minutes to a few hours.

5.15. Calculation of Short-wave Radio Circuits

Short-wave radio circuits are calculated in a way markedly different from that for very long, long, and medium waves, where one is interested in the requisite transmitter power and finds it by determining the signal strength at the receiver due to one kilowatt of radiated power.

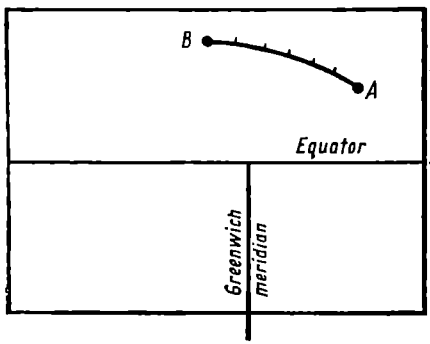


Fig. 5.33. Finding the MUF for a given propagation path

With short waves, one has to use frequencies not exceeding those maximally usable on a given propagation path and at a given state of ionization. At the same time, one has to reckon with radio-wave absorption which increases with decrease of frequency. So, the problem is one of finding optimum traffic frequencies.

It is convenient to break down the calculation procedure into two steps. In the first step, we shall find the optimum traffic frequencies and plot their schedule that will take into account the ionization of the *F2* layer at a particular hour of day, the season, and the phase in the eleven-year solar cycle. In the second step we shall determine the attenuation function for the frequencies thus found, so that the requisite transmitter power can be found for the desired field strength at the receiver, which is in turn dependent on the noise level and the desired operating reliability.

Step One. The easiest way is to use an ionospheric map. To begin with, an overlay is prepared on a sheet of tracing paper. Using the coordinate grid of the map (see Fig. 4.34), the equator, one of the meridians (say, the Greenwich one), and the terminal points of the radio circuit are drawn on the overlay. Now using a chart of great-circle paths, an arc of the great circle is drawn through the terminal points of the radio circuit on the overlay, and

a scale of distances is marked every 1000 kilometres, as shown in Fig. 5.33.

If the radio circuit is shorter than 4000 kilometres (which is the maximum skip distance for a wave reflected from the *F2* layer) there will be only one reflection, and the reflection point is suitably marked mid-way along the propagation path. Now the overlay is placed on a MUF-0 map for 0000 hours so that the equator and the meridian on the overlay register with those on the map, and the respective value of MUF-0 is read against the marked point at the centre of the route. This value is then entered under the second heading of Table 5.6. Let it be 6 megahertz.

Table 5.6. MUF DAILY SCHEDULE FOR RADIO CIRCUITS SHORTER THAN 4000 KILOMETRES

Hour of day	MUF-0, MHz	MUF-4000 MHz	MUF, MHz	OTF, MHz
0	6	12	11.3	9.6
2				
4				
.				
22				

Now the overlay is placed on MUF-0 maps for 0200, 0400, and 2200 hours, and the values found are entered under the same heading of the table. The same procedure is followed for MUF-4000 maps, and the next heading is filled in. Let the MUF-4000 for 0000 hours be 12 megahertz and the circuit length, 3000 kilometres. It is obvious that its MUF will be lower than that for the MUF-4000 but higher than that for the MUF-0. The requisite interpolation can be performed by use of the nomogram shown in Fig. 5.34 and a transparent ruler placed on the map so that it passes through the numeral *6* on the MUF-0 axis and the numeral *12* on the MUF-4000 axis. The position of the ruler is shown by the broken line in the figure. The point of intersection between the broken line and the vertical corresponding to the actual length of the radio circuit (in our case, 3000 kilometres) is then shifted onto one of the frequency scales, parallel with the slant lines. The result (11.3 MHz) is entered in the respective column of Table 5.6. The procedure is the same for all hours of the day. To be entered in the last column are the *optimum traffic frequencies* (OTF) which are 15 per cent lower than the MUF.

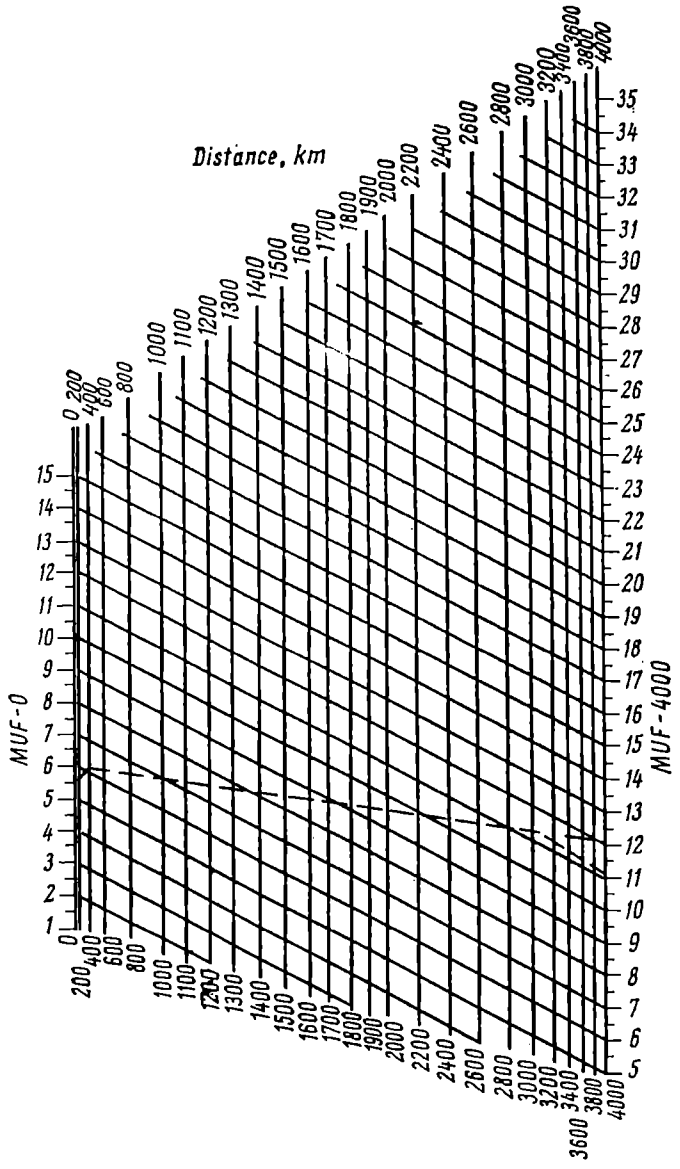


Fig. 5.34. Nomogram of MUF for the *F2* layer with propagation paths varying from 0 to 4000 km

The OTF is provided in order to ensure continuous working despite unforeseen fluctuations in ionosphere conditions.

If the radio circuit is longer than 4000 kilometres, two reflection points are marked on the route instead of one, each within 2000 kilometres of the respective terminal point. These are the points where oblique rays are reflected from the ionosphere. The calculation is performed using a single MUF-4000 ionospheric map. The overlay is placed on a 0000-hours map, and the values of the MUF-4000 are read off for the two reference points. The numerical results are entered in Table 5.7. The same procedure is followed for other hours of the day. The lowest of the two values of MUF-4000 is taken as the final MUF, because the conditions of reflection are decided by the "darker" reference point, that is, one with the lower value of the MUF-4000. The OTF is found in the same way as for a single reflection.

Table 5.7. MUF DAILY SCHEDULE FOR RADIO CIRCUITS LONGER THAN 4000 KILOMETRES

Hour of day	1st reference point, MUF-4000, MHz	2nd reference point, MUF-4000, MHz	MUF, MHz	OTF, MHz
0	10	14	10	8.5
2				
4				
.				
.				
.				
22				

The values from the last column in Tables 5.6 and 5.7 are then plotted to obtain a schedule of the OTF, such as shown in Fig. 5.35.

From reference to the MUF daily schedule we may conclude that operation on optimum traffic frequencies secures the least attenuation for radio waves in the ionosphere. However, it is hardly possible to re-tune the transmitter and receiver every two hours in actual service. Instead, use is made of a daily assortment of two or three frequencies. These recommended frequencies may be determined as follows. Using a particular daily MUF plot, one seeks to inscribe into it a two- or three-stepped curve (as shown in Fig. 5.36) such that the shaded areas are least in size. Another limitation is that the frequencies thus selected should be as close to the OTF as practicable.

In the example of Fig. 5.36, the daily assortment consists of three frequencies. The perpendiculars dropped from the steps mark on the

axis of abscissae the time for a change-over from one frequency to another—from a day-time through an intermediate to a night-time one. The ordinates of the respective steps give the recommended frequencies.

For radio circuits under simple conditions of illumination, that is, those with a single reflection point and also those with several

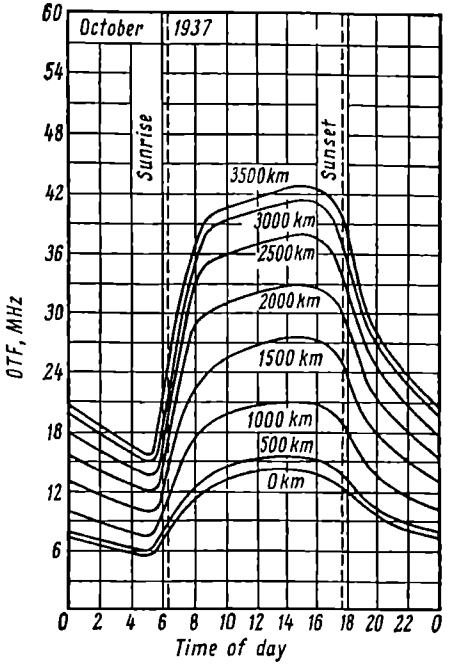


Fig. 5.35. Set of daily MUF plots for propagation paths differing in length

reflection points but all lying on the same meridian, a daily assortment of two or three frequencies is ordinarily sufficient to ensure continuous working on a round-the-clock basis. The conditions are more adverse for longer radio circuits oriented roughly along latitudes. On such circuits the western and eastern reflection points are illuminated very differently at dawn in winter, and the choice of frequencies is then an extremely tricky problem indeed. In fact, one has to use a daily set of four, and not of three, frequencies.

Step Two. The exact calculation of the attenuation function and, as a consequence, of the signal strength at the point of reception runs into difficulties which are aggravated by the lack of reliable

data about the absolute values of the absorption coefficient in the lower reaches of the ionosphere (the D and E regions).

Strictly speaking, in seeking a more or less exact value of the signal strength at the receiver one must know the trajectory of the wave. The unabsorbed field at the receiver, found in advance, is then multiplied by the factor $e^{-\int_s \delta dS}$, where δ is the absorption coefficient varying along the path, dS is the path element, and the

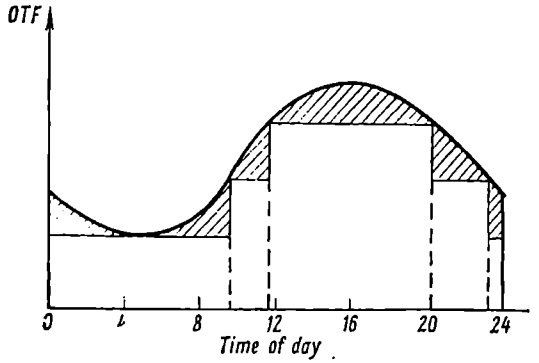


Fig. 5.36. Schedule of operating frequencies as found from a daily OTF plot

integral is taken over the path sections where the absorption due to the ionized layers is strongest. The same procedure is then applied to each of the rays experiencing different numbers of reflections, and the total field is found as the sum of the fields due to all the rays. A further difficulty is created by the need to allow for the effect of fading.

We shall discuss in detail the method due to Kazantsev first described in 1946 [90], [91] and then improved in 1956 [92].

The Kazantsev method is based on the assumption that in the general case the received field is produced by the combination of the rays that have experienced different numbers of reflections. The total field is given by

$$E_{rms} = \sqrt{\sum_{i=1}^{m} E_{rms(i)}^2} \text{ V/m} \qquad (5.26)$$

where m is the number of rays (never exceeding three).

It is to be noted that in most practical cases $E_{rms(1)} \gg E_{rms(2)} \gg E_{rms(3)}$, so that the calculations may be limited to the field of the ray that has experienced the least number of reflections.

The median value of the attenuation function for this ray is given by

$$F_{median} = \frac{1}{2}\left(\frac{1+R}{2}\right) R^{n-1} e^{-\sum_{j=1}^{n} \Gamma_j} \text{ mV/m} \qquad (5.27)$$

where R = modulus of the reflection coefficient for the earth's surface;
n = number of reflections from the ionosphere;
Γ_j = absorption coefficient at the jth point.

The number of reflections for a propagation path r kilometres long is given by

$$n \cong r_{km}/4000 \qquad (5.28)$$

The coefficient 1/2 in Eq. (5.27) represents an attenuation of 6 decibels in field strength (or power). Of this amount, Kazantsev

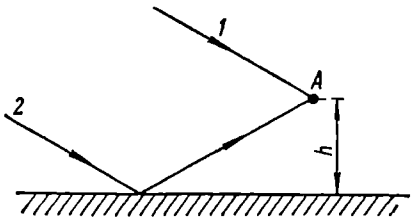


Fig. 5.37. Contribution of reflected waves to the total received field

ascribes 3 decibels to the fact that the receiving aerial is linearly polarized while the ray reflected from the ionosphere becomes elliptically (or even circularly) polarized. The remaining attenuation of 3 decibels is due to the losses caused by the magneto-ionic splitting of the ray [93].

The term $(1 + R)/2$ takes care of the effect of the ground-reflected wave at the receiving aerial. If the receiving aerial A (Fig. 5.37) were at an optimum height h above the ground, rays *1* and *2* would be in phase, and the total field (for horizontal polarization) would be given by $(1 + R)\,E_0$, where E_0 is the direct-ray field. In practice, the optimum conditions cannot be satisfied because of variations in the height of the reflecting layer, and so Kazantsev has introduced the arithmetic mean of the coefficient taking care of the effect of the ground-reflected wave, that is, $(1 + R)/2$. As a rule, R is unity very nearly, and so we may take it that $(1 + R)/2 \cong 1$. Kazantsev suggests a value of about 0.8 for R [93].

The third term takes into account the losses due to reflections from the ground at the intermediate points. For radio circuits with a single point of reflection this term vanishes, and $R^{n-1} = 1$.

The fourth term takes into account the absorption of radio waves in the ionosphere. The absorption coefficient at the jth point is given by

$$\Gamma_j = \Gamma_D + \Gamma_E + \Gamma_{F1} + \Gamma'_{F2} \qquad (5.29)$$

where Γ_D, Γ_E and Γ_{F1} designate the absorption as the wave travels both ways through the respective regions (so-called non-deviate absorption) and Γ_{F2} represents the loss upon reflection from the $F2$ layer (the deviate absorption). In cases where the wave is propagated owing to reflections from the E region, Eq. (5.29) takes the simpler form

$$\Gamma_j = \Gamma_D + \Gamma'_E \qquad (5.30)$$

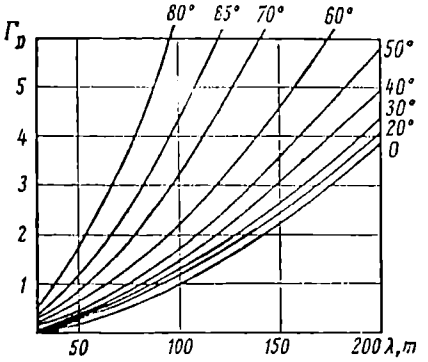


Fig. 5.38. Absorption coefficients for the D region as a function of wavelength for various angles of incidence φ_0; $\nu = 10^7$ sec^{-1}, $N = 10^3$ per cm^3

It appears that by day the wave is mainly absorbed in the D region. Since the D and E regions lie in a close juxtaposition and in view of the fact that the daylight ionization of the two is decided by the zenith angle of the sun, that is, obeys the same law, we may assume that the total absorption in both regions is proportional to the critical frequency of the E region. In other words, plots of the absorption coefficient related to the E region actually represent the attenuation in both the D and the E region.

The absorption coefficient is a function of three variables, namely the angle of incidence at the entrance to (the lower edge of) the ionized region, φ_0; the critical frequency of that region, f_c; and the operating frequency, f (or the respective wavelength).

Figs. 5.38 through 5.42 are plots for determining the absorption coefficients Γ_D, Γ_E, Γ'_E, and Γ'_{F2}. The operating wavelengths in metres are laid off as abscissa, and the sought values of the absorption coefficient as ordinate. The respective critical frequencies are given in the captions to the plots.

Kazantsev has shown that the attenuation in an ionized region is directly proportional to the maximum electron density which is in turn related to the critical frequency by Eq. (4.104), reproduced here for convenience

$$N_{max} = f_c^2/80.8 \text{ m}^{-3} \qquad (5.31)$$

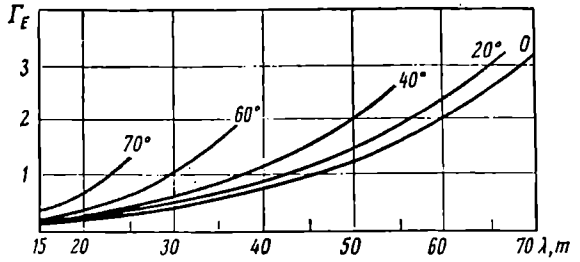


Fig. 5.39. Absorption coefficients for the E region as a function of wavelength at different angles of incidence φ_0; $f_c = 4 \times 10^6$ Hz

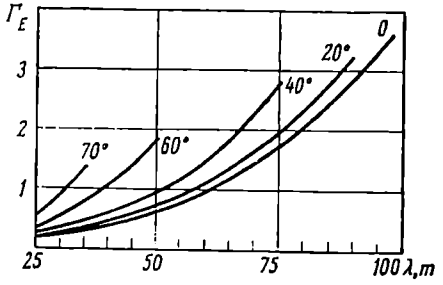


Fig. 5.40. Absorption coefficients for the E region as a function of wavelength at different angles of incidence φ_0; $f_c = 3 \times 10^6$ Hz

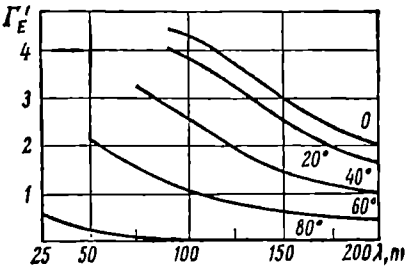


Fig. 5.41. Absorption coefficients for reflection from the E region as a function of wavelength at different angles of incidence φ_0; $f_c = 4 \times 10^6$ Hz

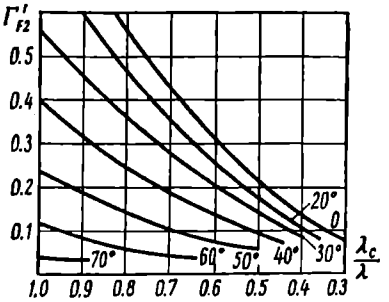


Fig. 5.42. Absorption coefficients for reflection from the $F2$ layer as a function of wavelength at different angles of incidence φ_0; $f_c = 5.7 \times 10^6$ Hz

On the basis of (5.31), the actual absorption coefficient can be found as

$$\Gamma = \Gamma_{fc0}\,(f_c/f_{c0})^2 \tag{5.32}$$

where f_c is the actual critical frequency, f_{c0} is the critical frequency for which the plot has been drawn, and Γ_{fc0} is the absorption as found from the plot.

Thus, if λ, φ_0 and f_c are known, we can find the absorption coefficients for all reflection points. The value of f_{cE} can be determined from an ionospheric map for the E region. The angle of incidence can be found from (4.126)

$$\tan \varphi_0 = \frac{a \sin \theta}{h_v + a\,(1 - \cos \theta)} \tag{5.33}$$

where the geocentric angle θ is given by

$$\theta^\circ = r_{\mathrm{km}}/(n \times 222.4) \tag{5.34}$$

Here r is the total length of the radio circuit, n is the number of reflections as given by (5.28), and h_v is the virtual height of the reflecting region (usually, *F2*) as found from a daily schedule of virtual heights.

To use the Kazantsev method practically, one has to have an ionospheric map for the absorbing regions, or a daily plot of the critical frequencies for those regions. The virtual heights involved are those where the MUF are reflected. The exact values of these heights can be found by means of the equivalence theorems (see § 4.12) and ionograms. For practical purposes, use may be made of averaged plots of daily variations in virtual heights.

The fact that the virtual heights h_v corresponding to the MUF are not maximum ones is clearly shown by the ionograms of Fig. 4.30. At frequencies approaching the MUF, reflections occur at two heights, one of which is greater and the other smaller than the reflecting height for the MUF.

To take into account the night-time attenuation, it should be assumed that $f_{cE} = 1$ megahertz throughout the night [93]. The effect of the eleven-year cycle is taken care of automatically, because the calculations are based on ionospheric predictions reflecting the actual ionization of the upper atmosphere.

Kazantsev's experiments [92] show that in a first approximation the attenuation-function values as obtained by (5.27) check well with the median values over periods of the order of hours for a given state of the ionized layers. Other investigators [33] have shown that the slow variations in field strength are described by the logarithmically normal distribution with a standard deviation of 8 decibels. The distribution function of such fluctuations is shown in

Fig. 3.34. Superimposed on the slow variations are the faster ones (fading) whose amplitude shows a Rayleigh distribution over ten-minute periods (§ 3.11). The distribution functions of fading for space-diversity reception with a varying number of aerials appear in Fig. 3.28.

An example involving the use of the Kazantsev method follows.

Example 5.4. Determine daily variations in the median field strength for $\lambda = 20$ m ($f = 15$ MHz) in July along the Tashkent Vladivostok radio circuit, assuming that the transmitter power is 40 kW and the transmitting aerial has a gain of $G_1 = 20$.

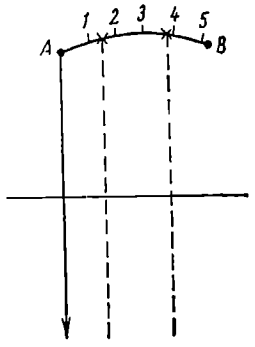


Fig. 5.43. Overlay with a plot of the propagation path. The crosses mark reflection points

Solution. To begin with, prepare an overlay (Fig. 5.43) as already explained. The total length of the radio circuit will be found to be 5100 kilometres. The reference meridian (the one from which the time count is to be kept) is drawn through point *A* (Tashkent).

The next step is to determine the number of reflections. For the radio circuit in question, there can be two reflections marked with the crosses on the overlay. The broken lines passing through the reflection points represent the local meridians.

By reference to a daily schedule of the MUF for the route, plotted in advance, the hours are determined when the specified frequency (15 megahertz) will not exceed the MUF. This gives the time interval during which continuous working is possible on the frequency chosen. For illustrative purposes, let this interval be between 0500 and 1900 hours (local Tashkent time).

The next step in the solution is based on averaged plots of variations in the critical frequencies and virtual heights for the absorbing and reflecting layers. In our problem, radio waves will be reflected from the *F2* layer (the daytime absorption in the *F2* layer may be neglected). On the other hand, as follows from Fig. 5.38, at wave lengths shorter than 50 metres the absorption in the *D* and *E* regions may be found from a single plot (Fig. 5.39). Thus, we should know (1) daily variation in the virtual height of the *F2* layer, and (2) daily variations in the critical frequency for the *E* region (see Fig. 5.44).

By placing the overlay on a coordinate grid, find the difference in longitude and the respective difference in time between the reflection points and Tashkent. The time differences are one and three hours, respectively.

The values of h_{vF2} and f_{cE} determined for the reflection points at local time are entered under headings 3, 5, 8, and 10 of Table 5.8.

The angles of incidence at the lower edge of the E region are found for each hour of day by Eq. (5.33), and the results are entered under headings 4 and 9 of Table 5.8.

The absorption coefficients for the E region at 15 megahertz (20 metres) and for the angles of incidence thus found are determined

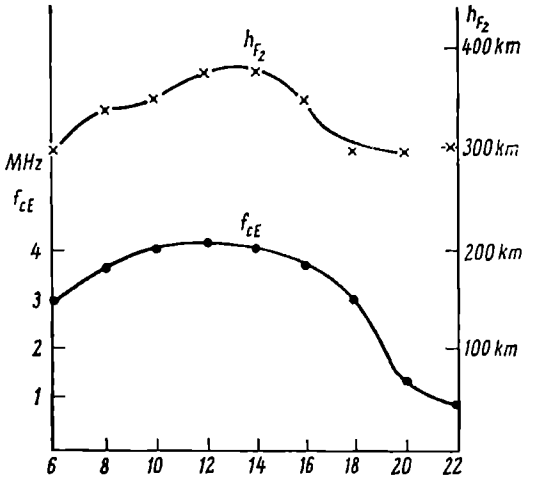


Fig. 5.44. Daily variations in virtual heights for the $F2$ layer and critical frequencies of the E region (to example 5.4)

by means of the plot of Fig. 5.39 for $f_{c0} = 4$ megahertz. The absorption coefficient for each hour of day is found from (5.32), and the results are entered under headings 6 and 11 of Table 5.8. Shown under heading 12 are the total absorption coefficients, while the median values of the attenuation function, F_{median}, are entered under heading 13. The median values of the rms field, E_{rms}, entered under the last heading 14 of the table, are found from

$$E_{rms\,(median)} = \frac{173\sqrt{P_{1\text{kW}}G_1}}{r_{\text{km}}} F_{median}\ \text{mV/m} \qquad (5.35)$$

In order to determine the transmitter power P_1 and the gains of the transmitting and receiving aerials G_1 and G_2 that will secure the desired operating reliability for a short-wave radio circuit, we must know:

(1) the median attenuation function;

(2) the distribution function of the slow variations in signal strength;

Table 5.8. FIELD CALCULATION BY THE KAZANTSEV METHOD

Hours	1st reflection point					2nd reflection point					$\sum \Gamma E$	E_{median}	E_{rms} μV/m
	local time	h_{vF2}, km	φ_0	f_{cE}, MHz	ΓE	local time	h_{vF2}, km	φ_0	f_{cE}, MHz	ΓE			
1	2	3	4	5	6	7	8	9	10	11	12	13	14
5	6	300	71°10′	3	0.39	8	340	69°40′	3.7	0.6	0.99	0.18	180
7	8	340	69°40′	3.7	0.6	10	350	69°10′	4.0	0.7	1.3	0.14	130
9	10	350	69°10′	4.0	0.7	12	370	68°30′	4.1	0.74	1.44	0.12	110
11	12	370	68°30′	4.1	0.74	14	370	68°30′	4.0	0.70	1.44	0.12	110
13	14	370	68°30′	4.0	0.70	16	350	69°10′	3.7	0.60	1.3	0.14	130
15	16	350	69°10′	3.7	0.6	18	300	71°10′	3.0	0.39	0.99	0.18	180
17	18	300	71°10′	3.0	0.39	20	300	71°10′	1.3	0.07	0.46	0.31	300
19	20	300	71°10′	1.3	0.07	22	300	71°10′	0.8	0	0.07	0.48	450

(3) the distribution function of the fast variations in signal strength due to fading for a varying number of aerials used in space-diversity reception;

(4) the number of aerials used for space-diversity reception;

(5) the safe signal-to-noise ratio for the expected form and quality of transmission;

(6) the bandwidth to be handled;

(7) the location of the reception point.

As is shown in Chapter 6, the noise level in the SW band is decided by atmospheric interference (assuming that there is no interference from other radio stations). A technique for determining the noise power P_n for a given location of the receiver and the specified bandwidth is described in detail in § 6.2. The requisite power at the receiver input is given by

$$P_2 = (S/N)\, P_n \text{ W} \tag{5.36}$$

The required transmitter power will then be found by Eq. (1.54) which is reproduced again for convenience

$$P_1 G_1 G_2 = 4\pi r P_2 / \lambda F \text{ W} \tag{5.37}$$

where F is the calculated attenuation function which provides for the compensation of both the slow and fast fluctuations in signal strength so as to secure the requisite reliability of transmission. It is given by

$$F_{\text{db}} = F_{median,\,\text{db}} + F_{1,\,\text{db}} + F_{2,\,\text{db}} \tag{5.38}$$

where F_{median} is found by the Kazantsev method, F_1 is found from the plot of Fig. 3.34 for a standard deviation of 8 decibels and the specified reliability, and F_2 is found from the plot of Fig. 3.28 for the specified reliability and the chosen number of aerials used in space-diversity reception.

If, as is often the case, the aerial gains G_1 and G_2 are specified in advance, the requisite transmitter power can be uniquely found by (5.37).

5.16. Propagation of Metric Waves Due to Reflections from the Regular Regions of the Ionosphere and the Sporadic E Layer

During years of medium solar activity the MUF is about 30 megahertz. When solar activity is high, there is a sharp increase in both the critical frequencies and the MUF. For example, in April, 1958 (a year with a sunspot number of 182.4, which is very near to a maximum) the *F2* layer at latitude 40° S reflected frequencies of about 50 megahertz ($\lambda = 6$ m).

High solar activity also entails an increase in the ionization of the Sporadic E layer so that it can now reflect frequencies up to 65 megahertz ($\lambda = 4.6$ m) near the equator.

To sum up, at the time of high solar activity the low-frequency end of the metric band may well be used for long-distance radio communication. The MUF and signal strength can be found by the methods described in connection with short waves.

5.17. Propagation of Metric Waves Through Scatter from the Ionosphere

The propagation of metric (VHF) waves through ionospheric scatter was discovered by Baily, et al. [94] in 1951.

In principle, the ionospheric scattering of metric waves does not differ from the scattering of ultra-short waves from local irregularities in the ionosphere, discussed in § 3.8. The only difference is

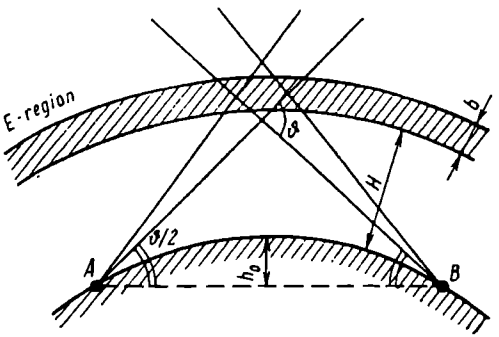


Fig. 5.45. Over-the-horizon propagation owing to ionospheric scatter

that in the metric band the scattering takes place in the D region (or in the lower part of the E region at night) due to local inhomogeneities in the electron density distribution. Diagrammatically, this is shown in Fig. 5.45.

As is seen, the mechanism involved is, in a sense, simpler than that discussed in § 3.8. For one thing, the volume V contributing to radiation scattering is more sharply defined. For another, the effective scattering cross-section per unit volume σ may be taken to be constant throughout the scattering volume.

As with tropospheric scattering, there are two problems involved. One is to derive an expression for the effective scattering cross section σ, and the other is to determine the volume contributing to radiation scattering.

From § 3.8 it follows that any one of two groups of equations may be used, according to the scattering model adopted. In one group the scattering parameter is $\overline{(\Delta\varepsilon'/\varepsilon')^2}$, and in the other, $\overline{\left(\frac{d\varepsilon'}{dh}-\frac{\bar{d}\varepsilon'}{dh}\right)^2}$ (Equations 3.45 and 3.51, respectively). With tropospheric scattering, the permittivity ε' and its fluctuations $\Delta\varepsilon'$ are independent of frequency. The situation is quite different with the ionospheric scattering of metric waves.

In § 4.17 it is shown that the permittivity of an ionized gas is given by

$$\varepsilon' = 1 - 80.8N/f^2 = 1 - f_0^2/f^2 \tag{5.39}$$

where

$$f_0 = 9\sqrt{N}\ \text{Hz} \tag{5.40}$$

is the so-called *plasma frequency* defined as the natural frequency of oscillation of a plasma of equal and oppositely charged species of particles.

Differentiating (5.39) yields

$$d\varepsilon' = -80.8\,dN/f^2 = -(dN/N)(f_0^2/f^2) \tag{5.41}$$

Dividing both sides into ε' gives

$$\frac{d\varepsilon'}{\varepsilon'} = -\frac{dN}{N}\frac{f_0^2}{f^2}\frac{1}{1-f_0^2/f^2} \tag{5.42}$$

Since in the metric band $f \gg f_0$, (5.42) takes a simpler form and the parameter $\overline{(\Delta\varepsilon'/\varepsilon')^2}$ may be written as

$$\overline{(\Delta N/N)^2}f_0^4/f^4 \tag{5.43}$$

In the other case, the scattering parameter takes the form

$$\overline{\left(\frac{d\varepsilon'}{dh}-\frac{\bar{d}\varepsilon'}{dh}\right)^2} \cong \frac{1}{N^2}\frac{f_0^4}{f^4}\overline{\left(\frac{dN}{dh}-\frac{\bar{d}N}{dh}\right)^2} \tag{5.44}$$

Equations (5.43) and (5.44) show that, in contrast to tropospheric scattering and irrespective of the scattering model chosen, the intensity of scattering in the ionosphere is strongly dependent on frequency—as the frequency is increased the scattering decreases. Because of this, ionospheric scattering can only be utilized in the metric band and then only on wavelengths longer than 5 metres.

Expressions for σ may be derived from (3.45) and (3.51) by substituting (5.43) and (5.44), respectively, for the scattering parameters.

The scattering angle entering these equations can be found from Fig. 5.45

$$\tan\frac{\Theta}{2} = \frac{h_0 + H}{AB/2} \tag{5.45}$$

The chord AB may always be replaced with the path length r as measured along the earth's surface. The height of the spherical segment is

$$h_0 \cong r^2/8a \text{ m} \tag{5.46}$$

and H is the average height of the scattering area, which may be set at 90 kilometres. Substituting in (5.46) and (5.45) gives

$$\tan \frac{\Theta}{2} = \frac{r^2 + 8Ha}{8ar} \tag{5.47}$$

As compared with tropospheric scattering, the ionospheric one occurs at a much greater height, because of which the signal may cover

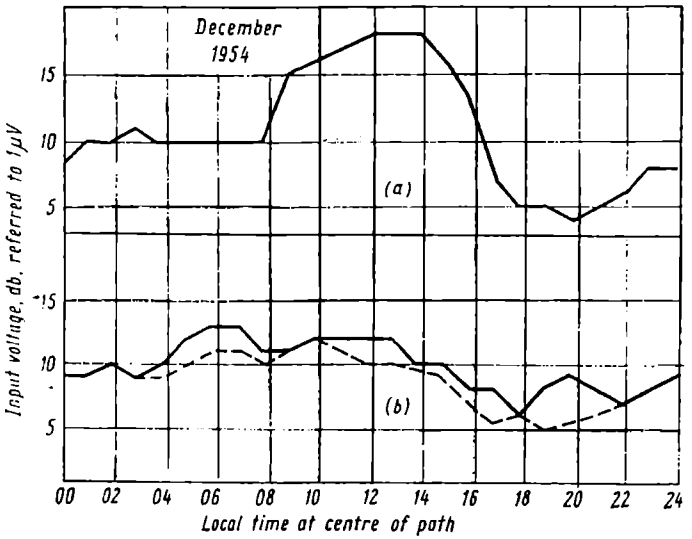


Fig. 5.46. Daily variations in signal strength: *a*—along a propagation path 1243 km long at $f = 49.8$ MHz; *b*—along a propagation path 2271 km long at $f = 36.0$ MHz

distances up to 2000 kilometres. However, at distances shorter than 800 kilometres the scattering angle Θ takes very high values, and the efficaciousness of ionospheric scattering is reduced.

The diurnal variations in signal strength with ionospheric scatter manifest themselves as an increase in the field strength by day and a more or less pronounced minimum at the centre of the path at 1900-2100 hours (local time). A typical plot of diurnal variations appears in Fig. 5.46

As with tropospheric scatter, the ionospheric-scatter signal is subject to both slow variations and fading, both described by about

the same distributions as in the former. It has been noted that, with other things being equal, the signal strength increases with geomagnetic latitude. The ionosphere disturbances accompanying global magnetic storms have no effect on the propagation of metric waves. In contrast, the auroral-zone and polar-cap absorption, when occurring to a considerable extent, may affect operation of ionospheric-scatter systems.

Fig. 5.47 relates the attenuation function to distance for five frequencies, as experimentally found for the ionospheric scattering

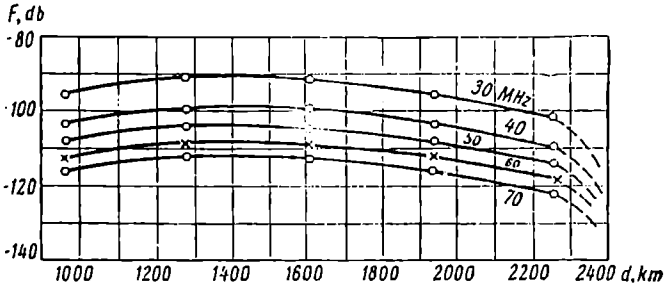


Fig. 5.47. Attenuation function plotted against distance for the propagation of metric waves at five frequencies by ionospheric scatter

of metric waves. The attenuation function corresponds to a scattered field level exceeded 99 per cent of the time [95]. The example that follows illustrates the use of the plot of Fig. 5.47.

Example 5.5. Determine the requisite transmitter power to ensure continuous working during 99 per cent of the time along a path 1800 km long at 60 MHz ($\lambda = 5$ m). The transmitting and receiving aerials are identical and have a gain of 30 db. The power received must be 5×10^{-15} W.

Solution: Using Eq. (3.78),

$$P_2 = \frac{P_1 G_1 G_2 \lambda^2 F^2}{(4\pi r)^2} \text{ W}$$

where F is the power attenuation function.

From the plot of Fig. 5.47,

$$F_{99\text{pc}} = -110 \text{ db},$$

whence

$$F = 10^{-(110/20)} = 10^{-5.5} = 3.16 \times 10^{-6}$$

Substituting the specified values of $G_1 = G_2 = 10^3$, $\lambda = 5$ m, and $r = 1.8 \times 10^6$ m, and the found value of F, we get

$$P_1 = 10.5 \times 10^3 \text{ W} = 10.5 \text{ kW}$$

The ionospheric scattering of metric waves can be utilized to advantage on distances up to 2000 kilometres, for transmission of signals with a bandwidth of up to six kilohertz. The transmission loss may be offset by an increase in transmitter power (up to tens of kilowatts) and by the use of highly directional aerials. Ionospheric-scatter systems are especially advantageous in the polar regions where global magnetic storms disrupt radio communication quite often while PCA events are a fairly rare occurrence.

5.18. Propagation of Metric Waves Due to Meteor Trails

In oscillograms of CW signals propagated by scattering from the *D* region one immediately notices spikes rising above the background of the uniformly fluctuating signals. As has been found, these spikes are produced by the signals reflected from meteor trails—shafts of ionized air which appear as a meteor enters the atmosphere, heats up and causes the surrounding gases to be ionized. The ionized gases form a trail behind the meteor. At first, a meteor trail is apparently not more than ten centimetres in diameter, but rapidly expands owing to molecular diffusion. In about a half-minute, turbulent diffusion leads to a still faster expansion of the trail. Convection currents and the wind twist the originally straight trail into fancy curves.

Every day, thousands of millions of such meteors hit the earth's atmosphere, producing ionized trails with an average extent of up to 25 kilometres [95] in the interval of heights from 80 to 120 kilometres; the average height of meteor trails is about 90 kilometres.

It may well be assumed that clearly discernible meteor trails are produced by particles weighing more than 10^{-5} gram and measuring more than 0.008 centimetre in radius each. The smaller meteors, so-called cosmic dust, produce the general background of ionization and, as already noted, may be one of the causes of irregularities in the *D* region.

A very important thing about this form of propagation is that the reflections of signals from newly formed trails are decidedly directional and look more like specular reflections than scattering. This is why some investigators prefer to speak of reflection rather than scattering of radio waves from meteor trails. Radio waves incident upon a meteor trail are mainly reflected in the direction for which the angle of reflection is equal to the angle of incidence. Therefore from among the multitude of meteor trails formed, only those suitably oriented may be utilized for radio communications. As the cross-section of trails increases, the diffusion grows more pronounced, and the scattered radiation loses its directional properties.

As meteor trails usually exist for 0.1 to 100 seconds and the trails capable of reflecting signals in the desired direction only occur from time to time, special techniques have had to be devised for transmission of information. In these "burst" communication systems, as meteor-scatter links have come to be known, the information is stored in advance in memory devices and then transmitted in bursts during the periods when suitable meteor trails can give a signal over the link.

These systems can only operate in the metric waveband. With increase of frequency the permittivity of meteor trails tends to unity and the irregularities in ionization they produce grow imperceptible.

A schematic presentation of a meteor-scatter link appears in Fig. 5.48 where A and B are the terminal points of the link and the heavy line represents a linear meteor trail. Q is a point on the meteor trail where the angle of incidence is equal to the angle of reflection (in the figure this is shown by the equality of the grazing angles δ). The sum of the distances, $r_1 + r_2$, to the point Q is a minimum. The broken line represents a projection of the meteor trail on plane AQB, and the angle β is the angle that the meteor trail makes with that plane. There may be two limiting cases. If the meteor trail is contained in plane AQB, the angle $\beta = 0$; if the meteor trail is at right angles to the plane, $\beta = 90°$.

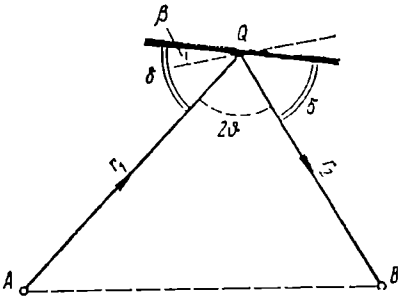


Fig. 5.48. Meteor-trail scattering

The signals reflected or scattered from meteor trails may be divided into two types according to the intensity of ionization in the trail—*underdense* and *overdense*.

With underdense trails, the plasma frequency

$$f_0 = 9\sqrt{N} \text{ Hz} \tag{5.48}$$

decided by the electron density is lower than the signal frequency, and so the relative permittivity

$$\varepsilon' = 1 - f_0^2/f^2 \tag{5.49}$$

does not reduce to zero. Under the circumstances, there is not sufficient ionization for reflection of the wave, and the returned signal may be calculated on the assumption of scattering from individual electrons.

In overdense trails the permittivity falls to zero in the centre of the trail and specular reflection of the signal can occur. The returned signal in this case is calculated on the assumption that the centre of the trail acts as a perfectly conducting (metallic) cylinder.

The theoretical consideration [96], [97] of the problem shows that for reflection of radio waves from underdense trails the attenuation function is given by

$$F = \frac{1}{2\pi}\left(\frac{\mu_0 e^2}{4m}\right)^2 \frac{r_1+r_2}{r_1 r_2} \frac{2\lambda q \sin^2\alpha}{(1-\cos^2\beta\sin^2\Theta)} \tag{5.50}$$

where

r_1 and r_2 = distances to point Q from transmitter and receiver;
β = angle between the axis of the trail and plane AQB;
Θ = half the angle AQB;
α = angle between the field vector and the direction of scattering, or forward scattering angle;
e and m = charge and mass of an electron;
μ_0 = permeability of free space;
q = linear electron density.

With an overdense trail, the attenuation function is given by

$$F = \frac{1}{2\pi}\left(\frac{\mu_0 e}{4m}\right)^{1/2}\left(\frac{\pi}{\exp}\right)^2 \frac{(r_1+r_2)\lambda\sin^2\alpha q^{1/2}}{r_1 r_2(1-\cos^2\beta\sin^2\Theta)} \tag{5.51}$$

where exp is the base of natural logarithms.

The signal returned from an underdense trail falls in power exponentially with time

$$P_2(t) = P_2(0)\exp(-t/\tau)\ \text{W} \tag{5.52}$$

where the time constant τ is given by

$$\tau = \frac{\lambda^2\sec^2\Theta}{32\pi^2 d}\ \text{s} \tag{5.53}$$

and d is a diffusion coefficient which depends on height. The value of d usually ranges between 1 and 10 square metres per second.

A plot of the signal strength as a function of time appears in Fig. 5.49.

With overdense trails, a different time relationship exists

$$P_2(t) = A\sqrt{td\ln\tau'/d}\ \text{W} \tag{5.54}$$

where A is a proportionality factor and the pulse duration τ' is given by

$$\tau' = \frac{10^{-7}e^2 q\lambda^2\sec^2\Theta}{4m\pi^2 d}\ \text{s} \tag{5.55}$$

It may be shown that the indeterminacy under the root sign reduces to zero when $t = 0$. It again reduces to zero when $t = \tau'$. When $t_0 = \tau'/\exp$, the radicand is a maximum. A plot of this time dependence is shown in Fig. 5.50.

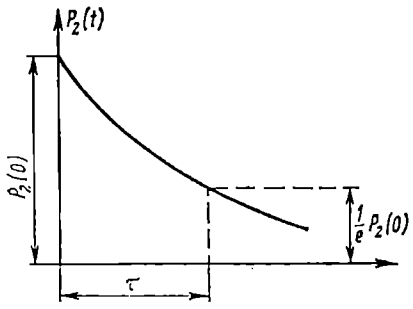


Fig. 5.49. Time plot of received power for reflection from underdense meteor trails

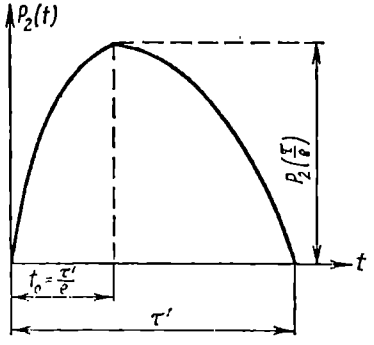


Fig. 5.50. Time plot of received power for reflection from overdense meteor trails

A comparison of Figs. 5.49 and 5.50 immediately shows that reflections from underdense and overdense trails produce considerably differing pulses.

In designing meteor burst communication systems, it is necessary to know the amplitude and duration of the high-level meteor signals. Just a few plots cannot of course cover the problem adequately, because these variables are functions of the location of the link, the transmitter power, the receiver gain and the mass of impinging particles. Yet the plots presented give some idea about the matter.

Figs. 5.51 and 5.52, taken from [97], relate the rate of occurrence to a specified relative amplitude and a specified duration of meteor signals. The broken lines in Fig. 5.51 apply to underdense and overdense meteor trails. As will be noted, there is a drastic decrease in the number of meteors with increase of the amplitude and duration of the ionized trails. It is also seen that weak returned signals of short duration have a greater probability of occurrence.

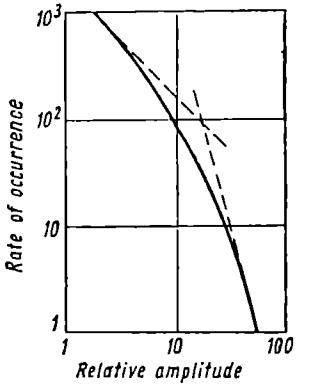


Fig. 5.51. Rate of occurrence related to a specified relative amplitude of meteor trails

From the view-point of radio communications, the most important characteristic is the duration of suitably oriented meteor trails producing reflections of a strength sufficient for the transmission of information. This duration is again a function of the transmitter power and the receiver sensitivity. As the transmission loss $L = P_1/P_2$ (which may be sustained by the choice of a suitable equipment for the meteor-scatter link) increases, the duration of transmission also increases. An idea about the situation can be had from the plot of Fig. 5.53 taken from [97], where the local time t is laid off as

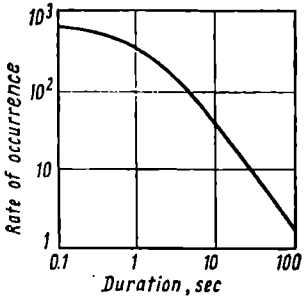


Fig. 5.52. Rate of occurrence related to a specified duration of meteor trails

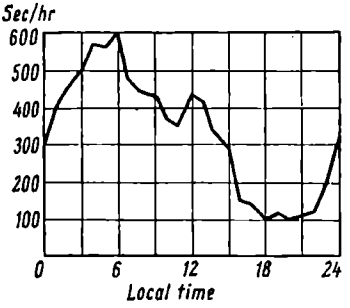


Fig. 5.53. Duty cycle of a meteor burst communication link as a function of the time of day

abscissa and the number of seconds during each hour when the transmission of information is possible, as ordinate. The plot has been constructed for a meteor-scatter link about 1000 kilometres long and operating on 40 megahertz. The duty cycle of this link is between 16 and 3 per cent (at 0600 and 1800 hours, respectively). The duty cycle is the ratio of the time that communication is possible to the total time.

A major advantage of meteor-scatter links over those using ionospheric scattering is that they do well with lower transmitter power (under 500 watts) and simpler and less expensive aerials.

Meteor burst communication systems may cover distances up to 2000 kilometres with a bandwidth of 3 kilohertz. The relatively low power of the transmitters is more than offset by the use of more sophisticated transmitting and receiving equipment incorporating memory devices and automatic controls. Meteor bursts can be used on frequencies from 30 to 50 megahertz.

5.19. Modes of Propagation for Metric Waves

The metric waveband is probably the only band where all modes of propagation are possible. In the three preceding sections we have examined the long-distance propagation of the longer metric waves

(4 to 10 metres) by way of the ionosphere. They can also propagate as ground and tropospheric waves along the earth's surface. Last but not least, metric waves are capable of passing through the ionized layers and being propagated as direct waves, notably in space communication.

A detailed discussion of the propagation of ground and tropospheric waves will be found in Chapters 2 and 3.

5.20. Propagation of Decimetric and Centimetric Waves

Radio waves in the decimetric (UHF) and centimetric (SHF) bands cannot be reflected from the ionosphere, and so they cannot be propagated as ionospheric waves. Over short distances (within the line of sight) they are propagated as ground waves (it is relevant to recall that diffraction is negligible at these high frequencies). Long-distance propagation is secured by the tropospheric mode, mainly due to scattering from irregularities in the troposphere and less due to ducted modes in the troposphere.

Radio waves in these bands are refracted in the ionosphere very little if at all and readily pass through it. This is why they can be propagated as direct waves and find use in space communication.

Decimetric (UHF) waves do not practically experience molecular absorption or absorption in precipitation particles. However, the absorption by rain, hail, and snow is perceptible at wavelengths shorter than three to five centimetres. The absorption by water vapour, that is, molecular absorption, is noticeable at a wavelength of 1.35 centimetres, that is, at the lower boundary of the SHF band. Obviously, the molecular absorption of centimetric waves may likewise be neglected.

5.21. Propagation of Millimetric Waves

Naturally, the ionosphere has no effect on the propagation of millimetric waves. In contrast, the troposphere causes the bending (atmospheric refraction) of millimetric waves.

Rain, fog, hail, snow and other forms of precipitation particles are responsible for a marked absorption in the millimetric band. In fact, a heavy rain or a dense fog will completely stop millimetric waves. There is also a strong molecular absorption by the tropospheric gases, especially water vapour and oxygen, irrespective of the presence or absence of precipitation particles.

Referring to Fig. 3.39 which relates the absorption coefficient (in decibels per kilometre) to frequency, there are four "windows" in the wavelength range 1-10 millimetres, where the absorption is relatively insignificant, namely around 1.2 millimetres ($\delta =$

= 0.7 decibel per kilometre), 2 millimetres (δ = 0.3 decibel per kilometre), 3 millimetres (δ = 0.22 decibel per kilometre) and at 8.6 millimetres with an absorption of 0.07 decibel per kilometre. The fourth "window" was widely used for radar (the so-called K_A subband) during the second world war. It is to be recalled that the plot of Fig. 3.39 holds for a liquid water content of 7.75 grams per cubic centimetre. At an ambient temperature of 15°C this corresponds to 60 per cent relative humidity. As follows from Table 3.6, at 23°C the liquid water content may be 23 grams per cubic centimetre which corresponds to a three-fold increase in the absorption

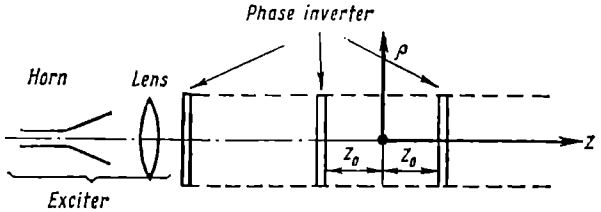


Fig. 5.54. A beam waveguide

coefficient. The same increase will occur in the minimum values of absorption. For example, in the 3-mm subband the absorption will be 0.7 decibel per kilometre. For a path length of 50 kilometres (which is a standard hop length for microwave radio relay links) the total attenuation will be 35 decibels, a rather sizeable loss.

The fact that millimetric waves are strongly attenuated owing to molecular absorption and are completely stopped by rain and fog seriously handicaps their use for ground-based communication systems, that is, ones utilizing ground waves. On the other hand, millimetric waves are advantageous in space communication outside the troposphere with no thunderstorms or water vapour to interfere.

A more recent trend has been towards using millimetric waves in conjunction with tubular waveguides propagating an H_{01} mode whose absorption, in contrast to other modes, decreases with increase of frequency. According to [98], waveguide communication systems are expected to use wavelengths between 8 and 4 millimetres.

There is also Goubau's idea [99], [100] to use "beam" waveguides for the millimetric waveband. "Beam" waveguides ensure a high concentration of energy in a chosen direction of propagation by means of a system of electron-optical lenses arranged at suitable intervals.

For its operation a beam waveguide (Fig. 5.54) depends on the fact that a cylindrical wave propagated in a specified direction may be represented as a collection of inhomogeneous plane waves [101], that is, waves whose planes of constant phase are at right angles to the direction of propagation and the amplitudes in those planes

have a certain distribution. Goubau has shown that, by use of suitable phase transformers disposed at equal intervals, it is possible to isolate several waves of the lowest mode from among the multiple of those being propagated, such that the energy distribution relative

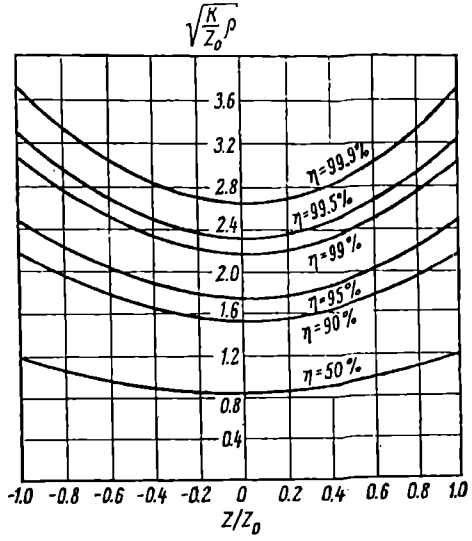


Fig. 5.55. Radius ρ as a function of distance z

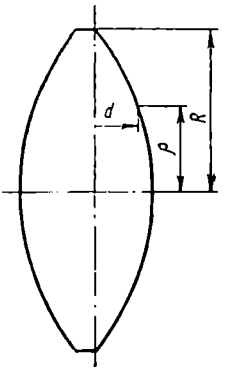


Fig. 5.56. Half-thickness d of a lens as a function of radius ρ

to the waveguide axis is fairly compact. Such waves show a fast decrease in amplitude in the radial direction of the cross-sectional plane.

Fig. 5.55 is a plot of the relation between the radius ρ which encloses the specified share of the total energy being propagated (50 to 99.9 per cent), and the distance z. According to Fig. 5.54, $z = 0$ corresponds to a point at the middle of the segment, and $z = \pm z_0$ locates the positions of adjacent lenses. The ratio z/z_0 is laid off as abscissa, and $\rho\sqrt{k/z_0}$ (where $k = 2\pi/\lambda$ m^{-1}) as ordinate.

The radius ρ is a minimum in the middle of the segment and is a maximum at its ends. Fig. 5.55 represents a steady-state case occurring, as Goubau believes, after ten to twenty transformations. In his experiments which were carried out on $\lambda = 1.3$ centimetres. Goubau spaced phase transformers $2z_0 =$ one metre apart. From the plot of Fig. 5.55 it is an easy matter to determine that for $z = z_0$ the radius ρ containing 99 per cent of all energy is 13.6 centimetres.

A phase transformer should be such that

$$\psi(\rho) = \rho^2 k/2z_0 \qquad (5.56)$$

where ψ is the phase shift that a wave experiences as it passes through the transformer at a distance ρ from the centre of the transformer

Eq. (5.56) is satisfied by a bi-convex lens (Fig. 5.56) fabricated from a dielectric of permittivity ε'. The half-thickness d of the lens must be such that

$$d = \frac{R^2 - \rho^2}{4z_0(\sqrt{\varepsilon'} - 1)} \text{ m} \qquad (5.57)$$

where R is the radius of the lens.

Owing to the high concentration of energy along the axis of the beam waveguide, the diffraction losses, that is, the losses caused by the natural dispersion of waves generated by a source with a finite aperture, are very low. According to Goubau, the total loss per interval was as low as 0.052 decibel at $f = 35$ gigahertz ($\lambda = 0.86$ millimetre) as measured under steady-state conditions.

5.22. Propagation of Optical-frequency Radio Waves

Optical-frequency radio waves (arbitrarily ranging between 0.03 micron and 2 millimetres in wavelength) can be propagated only as ground and direct waves. Atmospheric refraction causes the bending of the ray path. Owing to the fact that the polar molecules of water vapour have a lesser effect at these frequencies, the radius of curvature for normal atmospheric refraction and oblique rays is about 50,000 kilometres as against 25,000 kilometres for ultra-short waves.

A heavy rain or a dense fog will completely stop the passage of optical-frequency radio waves through the troposphere. Haze and overcast absorb them to a point where the range of communication may be cut down to a few kilometres.

With no atmospheric precipitation present, communication with optical-frequency radio waves is possible only in the wavelength interval between 0.4 and 20 microns, and then only within the limits of the windows listed in Table 3.7. This is why it will pay to use optical-frequency radio waves only in localities where atmospheric precipitations are a rare occurrence. However, they are well suited for space communication outside the troposphere.

As is noted in § 3.15 when optical-frequency radio waves are intended for use in a ground-based communication system it is possible, at least in principle, to enclose the entire propagation path in a pipe shutting atmospheric precipitation out of the propagation medium. Of course, this pipe is not a waveguide and only serves as a protection against precipitation. The use of such an enclosure is favoured by the high concentration of energy in the beam and the high directivity of the aerials used.

Chapter Six

Atmospheric and Cosmic Noise

6.1. Main Sources of Noise

It has been noted more than once in the previous chapters that the conditions of reception are determined not so much by the strength of the signal at the receiver as by its ratio to noise strength. This is why it is important to know the noise level. Once the signal-to-noise ratio is known, one can readily find the requisite signal strength at the input to the receiver and, as a consequence, all other characteristics of the radio circuit.

The noise power present at the input to a receiver is the sum of internal (receiver) noise and external noise. Leaving out internal noise (treated in texts on radio receivers), we shall be concerned with external noise.

The main types of external noise are man-made interference, atmospheric noise, cosmic noise, and also thermal noise, that is, one produced by the heated atmosphere and earth surface.

Man-made interference comes from electrical motors, home appliances, induction heaters, automobile ignition, and other industrial and household electrical plant. It is especially strong in large cities.

The best method to reduce the level of man-made noise is to suppress it at the source by use of filters and screens which will prevent the emission of radio waves, on one hand, and the propagation of radio waves in the supply wires, on the other. The level of man-made interference can be measured by suitable instruments which can also locate the offending equipment and check on the efficiency of the measures taken.

6.2. Atmospheric Noise

Atmospheric noise, also known as precipitation static and atmospherics (abbreviated to sferics), comes from naturally occurring electrical discharges, such as lightning flashes.

Oscillograms of lightning flashes show that such a discharge consists of (1) a predischarge, an initial stroke during which a pilot

streamer slowly moves from the cloud towards the ground, followed by a leader with a current of 100 to 300 amperes, moving in a series of steps of 25 to 30 microseconds; and (2) the main discharge, a return stroke with a peak current of 10 to 100 kiloamperes, progressing from the earth to the cloud through the path blazed by the stepped leader, as shown in Fig. 6.1 [102].

The leader emits a continuous spectrum with a peak in the range 30-50 kilohertz, having a spectral distribution such that the noise power is inversely proportional to frequency ($1/f$ noise). The return stroke, lasting about 100 microseconds, also emits a continuous spectrum with a peak in the range 4-8 kilohertz, having a spectral distribution such that the noise power is inversely proportional to the square of frequency ($1/f^2$ noise). These and the resultant spectra at a distance of about 2 kilometres from a lightning discharge are shown in Fig. 6.2. As is seen, lightning discharges emit radio waves within a fairly broad range of frequencies — from extremely low to tens of megahertz (Fig. 6.2 shows only a part of the entire spectrum).

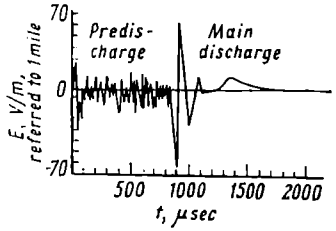


Fig. 6.1. Time plot of the field due to a lightning discharge

Atmospheric noise may be caused by local thunderstorms and by global sources. This is why atmospheric noise may affect operation of a receiver even when there is no thunderstorm going on nearby. When atmospheric noise is caused by local thunderstorms its power in the frequency range $f > 8$ kilohertz will be inversely proportional to frequency ($1/f$), according to the plot of Fig. 6.2. When atmospheric noise is caused by distant lightning discharges, the noise power will on the whole be decided not only by the spectrum of Fig. 6.2 but also by the conditions under which the respective frequencies are propagated from the source of lightning discharges to the receiver. Under the circumstances, each lightning discharge ought to be treated as a source of radio waves with a continuous frequency spectrum, propagated to the receiver exactly as the waves emitted by radio aerials do.

About a hundred lightning discharges occur every second on the earth. Overlapping, they are perceived by the human ear as continuous rustling sounds. Most of all, lightning discharges occur in the tropical belts of the continents, with the zone where they take place moving along the equator together with local noon.

The 8th Plenary Session of the CCIR, held in Warsaw in 1956, recommended special maps for determining the noise level, compiled separately for winter (December, January, and February), spring,

summer, and autumn and for six 4-hour intervals daily (0000-0400, 0400-0800 hours, etc.). The complete set includes twenty maps [33], because some have been combined. An example of such a map is shown in Fig. 6.3. It uses the same projection as ionospheric maps. Each map shows isopleths of noise level for a frequency of one megahertz, expressed in decibels relative to kTB (watts), where k is

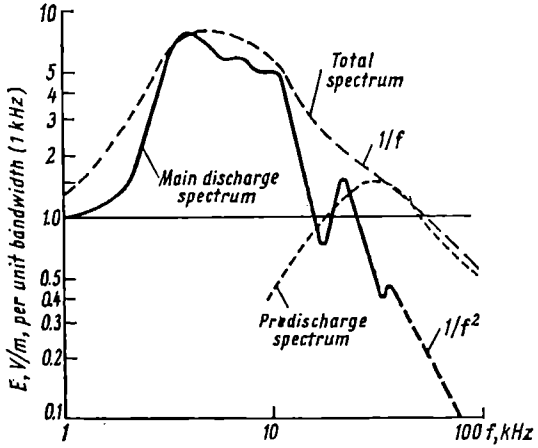


Fig. 6.2. Frequency spectrum of a lightning discharge

Boltzmann's constant (in joule/degrees), T is the absolute temperature assumed to be 288° K, and B is the effective bandwidth of the receiver (in hertz). The numerals on the isopleths give the median noise level for the respective 4-hour interval.

It is convenient to express the noise power in kTB units because the noise power at the input to a receiver is also expressed that way

$$P_n = NkTB \text{ W} \tag{6.1}$$

where N is the noise factor of the receiver.

The values of n indicated on a map can be converted to the atmospheric noise power available at the input to a receiver as follows.

The product nT may be thought of as the temperature of equivalent black-body radiation that would produce the same power flux density normal to its surface at frequency f and in bandwidth B as the atmospherics have produced at the receiver input.

A black body raised to T_e produced a power flux density given by the Rayleigh-Jeans equation

$$S = \frac{8\pi k T_e f^2 B}{c^2} \text{ W/m}^2 \tag{6.2}$$

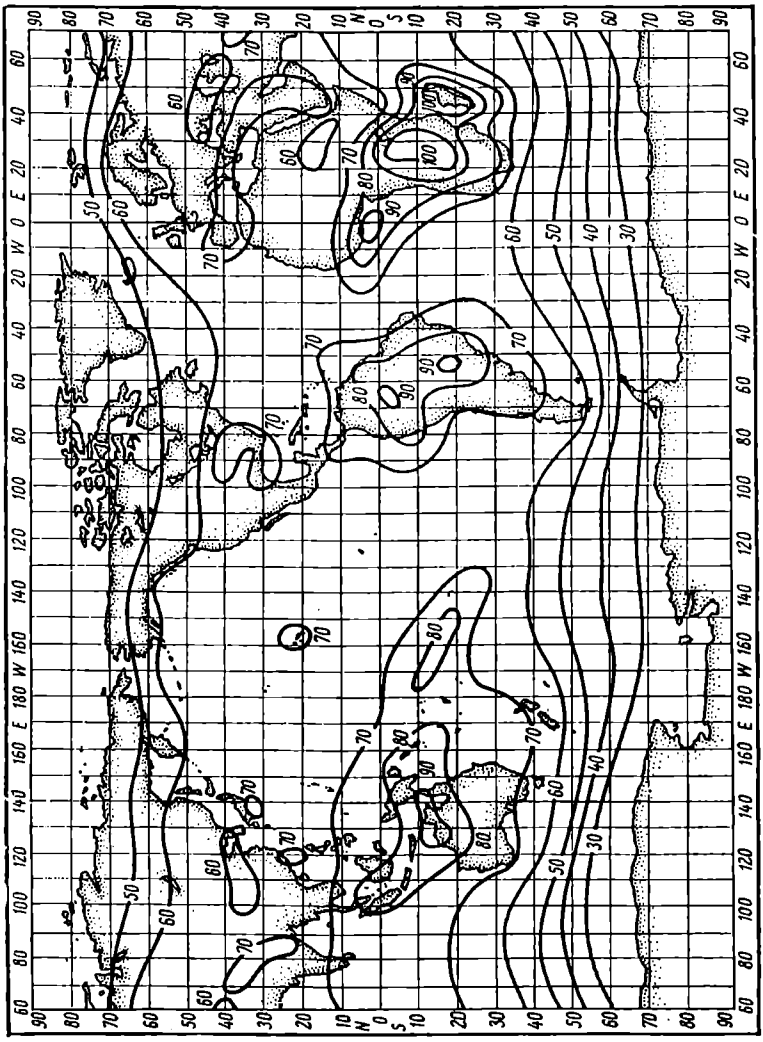


Fig. 6.3. Isopleths of atmospheric noise for a frequency of 1 MHz in decibels referred to kTB

Eq. (6.2) describes the flux density of an unpolarized radiation. Since, however, receiving aerials respond solely to plane-polarized waves, we write

$$S' = \frac{4\pi k T_e f^2 B}{c^2} \text{ W/m}^2 \tag{6.2a}$$

Recalling that

$$S' = E_{rms}^2/120\pi \text{ W/m}^2 \tag{6.3}$$

equating (6.2a) and (6.3), and substituting nT for T_e, we get

$$E_{rms} = \frac{4\pi}{\lambda}\sqrt{30nkTB} \text{ V/m} \tag{6.4}$$

In a vertical receiving aerial with an effective height h_e, the vertical component of the wave field induces an emf given by

$$v = E_{rms} h_e \text{ V} \tag{6.5}$$

If the input impedance of a receiver is purely resistive, R, the noise power at the receiver input is given by

$$P_n = v^2/R = E_{rms}^2 h_e^2/R = \frac{16\pi^2 30 h_e^2}{\lambda^2 R} nkTB \text{ W} \tag{6.6}$$

It is not difficult to determine the effective height of the receiving aerial for which the noise power will be given by

$$P_n = nkTB \text{ W} \tag{6.7}$$

From the condition $\frac{16\pi^2 30 h_e^2}{\lambda^2 R} = 1$ we find that

$$h_e = \lambda\sqrt{R}/69 \text{ m} \tag{6.8}$$

If the input resistance is, as usual, $R = 70$ ohms, then

$$h_e = 0.121\lambda \text{ m} \tag{6.7a}$$

The map of Fig. 6.3 has been compiled for $f = 1$ megahertz ($\lambda = 300$ metres). Therefore, for a vertical aerial with an effective height $h_e \cong 36$ metres the noise power will be n decibels above kTB. For $B = 1$ kilohertz,

$$kTB = 1.38 \times 10^{-23} \times 288 \times 10^3 \cong 4 \times 10^{-18} \text{ W}$$

In order to find the noise power for frequencies other than one megahertz, use should be made of another series of plots one of which appears in Fig. 6.4. The ordinates for $f = 1$ megahertz (laid off as abscissa) give the decibel values of n from 0 to 100.

If the map of Fig. 6.3 gives, say, a noise level of 60 decibels, we identify the respective curve on the plot of Fig. 6.4 and find the value of n for any frequency from 0.01 megahertz (that is,

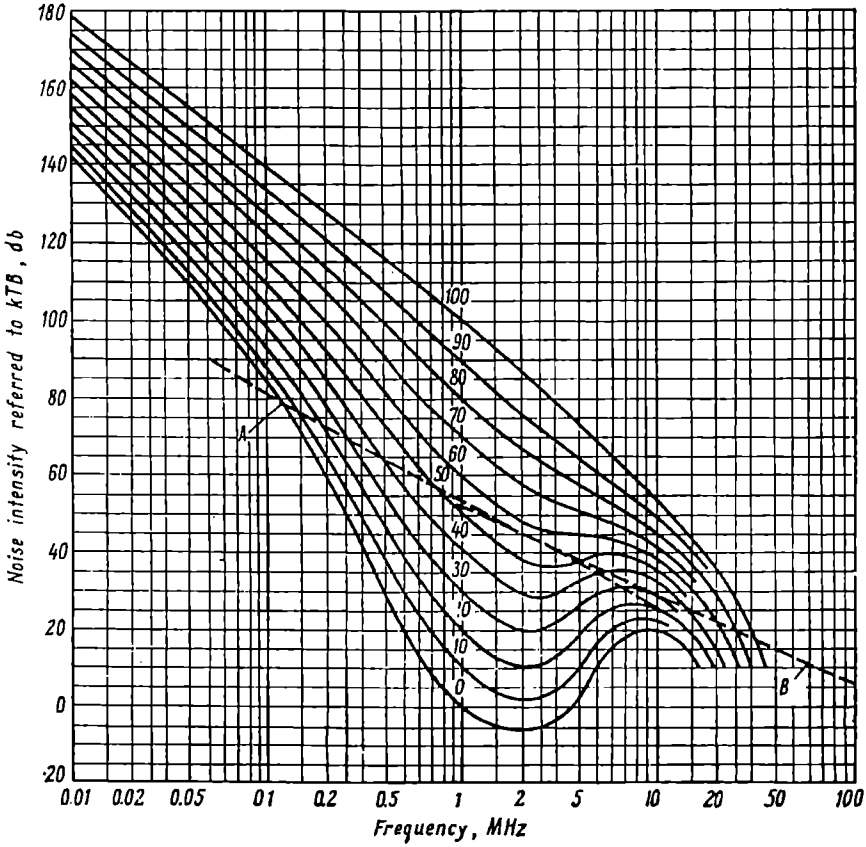


Fig. 6.4. Frequency plot of atmospheric noise (in decibels referred to kTB)

10 kilohertz) to 50 megahertz. The shape of the curves in Fig. 6.4 is explained by the conditions of propagation for the respective frequencies, while the sudden decrease in the noise level at frequencies exceeding 20-30 megahertz is explained by the fact that these frequencies pass through the ionosphere.

The line A in the plot of Fig. 6.4 represents the approximate level of man-made noise (for areas where this type of interference is low), and the line B represents the level of cosmic noise. As is seen, the overall noise level in the metric waveband is decided by cosmic noise.

In order to convert the values of n as found from the map of Fig. 6.3 and the plot of Fig. 6.4 into field strength, use may be made of Eq. (6.4). Writing the respective frequency instead of the

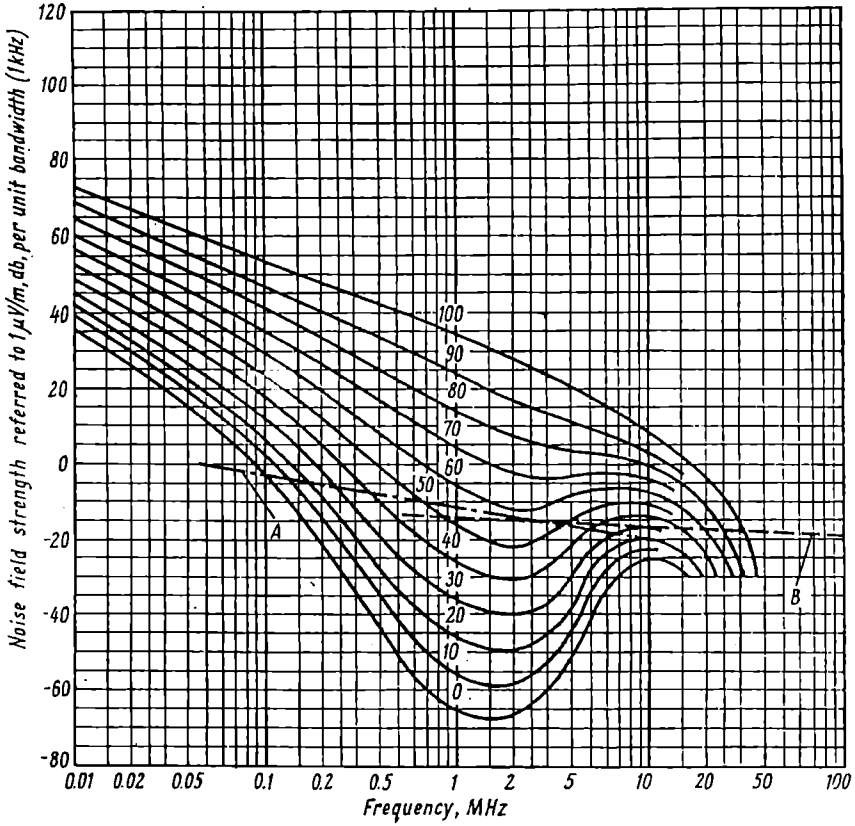


Fig. 6.5. Frequency plot of atmospheric noise (in decibels referred to 1 μV/m)

wavelength, using the numerical values of Boltzmann's constant ($k = 1.38 \times 10^{-23}$ joule/degree) and absolute temperature ($T = 288°$ K) and setting $B = 10^3$ hertz, we may write (6.4) as

$$E_{rms} = 4.55 \times 10^{-4} f_{\text{MHz}} \sqrt{n} \ \mu\text{V/m} \qquad (6.9)$$

For practical purposes, the level of atmospheric noise in terms of equivalent field strength may be found by use of the plot appearing in Fig. 6.5 which is similar to that of Fig. 6.4 except that the equivalent field strength in decibels, referred to one microvolt per metre for $B = 1$ kilohertz, is laid off as abscissa.

It is relevant to note that the field strength due to atmospheric noise is the rms value of the signal field that would develop the same power at the receiver input as the atmospheric noise. This

is the reason why the noise field is proportional to the square root of the bandwidth.

Example 6.1. Using the map of Fig. 6.3 and the plot of Fig. 6.4, determine the noise field strength near Leningrad (longitude 30°E, latitude 60°N) for a frequency of 0.1 megahertz and a bandwidth of 3 kilohertz.

Solution: Referring to the map, for the Leningrad area $n = 60$ decibels $= 10^6$ (the conversion formula is $n_{db} = 10 \log_{10} n$ because kTB has the dimensions of power).

Referring to the plot of Fig. 6.5, $n = 60$ decibels at $f = 0.1$ megahertz corresponds to $E_{rms} = 30$ decibels relative to one microvolt per metre.

Noting that $30 = 20 \log_{10} E_{rms}$, we get

$$E_{rms} \cong 32 \ \mu V/m$$

If necessary, it is possible to find the noise power at the receiver input for fixed values of h_e and R by use of Eq. (6.6).

6.3. Cosmic Noise

As we have seen, at frequencies exceeding 30 megahertz the level of atmospheric noise is drastically reduced because the associated waves (atmospherics) cannot be propagated by the ionospheric mode. Therefore in the metric band the main source of interference is the radio emission coming from various radio sources in and outside the Galaxy and from the sun. This is cosmic noise.

The distribution of radio sources in the sky has been investigated rather well. Detailed maps have been compiled, bearing isophots, or lines of equal emission rate. It has been found that galactic radio emission is extremely steady in magnitude. The variations in the emission rate observed have been found to be caused by variable absorption in the ionosphere. This discovery has provided a means for ionospheric studies—so-called *riometers*, which are metric-wave-band receivers registering the intensity of cosmic radio emission. While ionospheric stations illuminate the ionosphere and note its response in the form of reflected signals, riometers are passive devices picking up the external radiation that happens to pass through the ionosphere. Sudden changes occur only in solar radio emission, especially during disturbances.

It is customary to express the intensity of cosmic noise in terms of either the temperature T_e (degrees Kelvin) of equivalent black-body radiation, or power flux density (watts per square metre), or the rms value of the equivalent field strength (microvolts per metre). The three quantities are related by Eqs. (6.2a) and (6.4).

Using the numerical values of Boltzmann's constant and the velocity of light, these equations may be re-written as follows

$$S' = 1.9 \times 10^{-24} f^2_{\text{MHz}} T_e B_{\text{kHz}} \text{ W/m}^2 \quad (6.10)$$

$$E_{rms} = 2.68 \times 10^{-5} f_{\text{MHz}} \sqrt{T_e B_{\text{kHz}}} \text{ μV/m} \quad (6.11)$$

A typical record of one week's measurements of the cosmic-noise field strength on frequencies of 25, 50, 75, and 110 megahertz appears

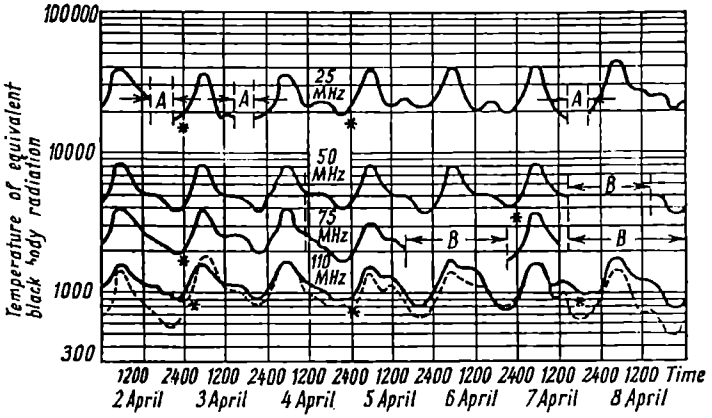


Fig. 6.6. Typical record of one week's measurements of the cosmic-noise field strength on frequencies of 25, 50, 75, and 110 MHz

in Fig. 6.6 [103]. As is seen, the noise field strength is drastically reduced with increase of frequency.

The cause of diurnal variations in the galactic noise power lies in the fact that the radio sources in outer space vary in intensity. As the earth rotates about its axis, each point on its surface comes under the differing radio sources with a periodicity of twenty-four hours.

The absolute value of normal cosmic noise can be determine. from the plot of Fig. 6.7 (curves *A*) [103]. The top curve represents the normal daily maxima, and the bottom curve, the normal daily minima. For comparison, there are also the maximum and minimum values of atmospheric noise (at *D*). The curve *B* of the same plot represents the noise intensity of well-designed valve receivers, while the curve *C* does so for the noise intensity produced by black-body radiation at the temperature of the earth ($T = 300°\text{K}$). In all cases, the noise intensities are expressed in microvolts per metre for one-kilohertz bandwidth of the receiver.

The plot of Fig. 6.7 displays the fact that cosmic radio noise is decisive in the frequency range from 20 to 200 megahertz. At lower

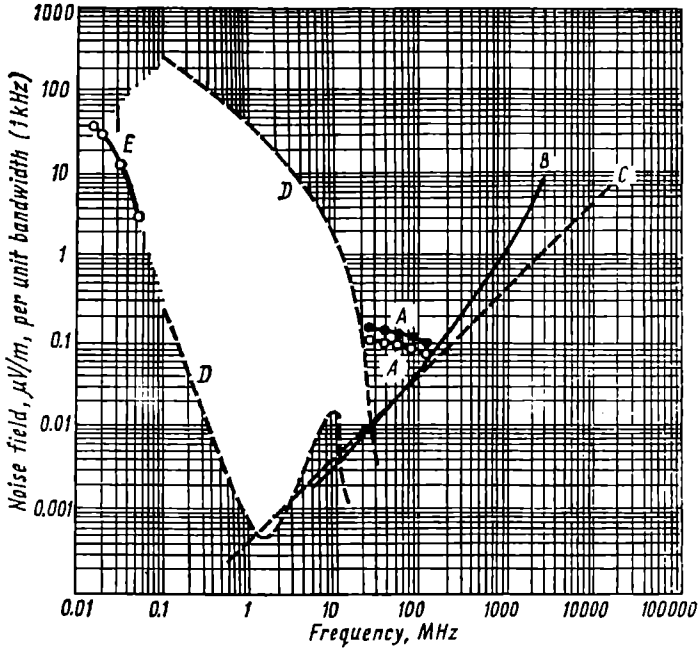


Fig. 6.7. Cosmic-noise field plotted against frequency

frequencies the dominant role is played by atmospherics, while at higher frequencies this is true of internal (receiver) noise.

The dependence of the maximum and minimum noise intensities on frequency is shown with greater detail in Fig. 6.8 [103] where frequencies are plotted as abscissa, while the temperature of equivalent "black-body" radiation, the power flux density in watts per square metre per unit bandwidth (one kilohertz), and the equivalent field strength in microvolts per metre for a bandwidth of one kilohertz, as ordinate. It may be seen that when the intensities are expressed in terms of temperatures of equivalent black-body radiation, they vary inversely as the 2.3 power of the frequency, when in terms of electric field strength, inversely as the 0.15 power, and when in terms of power flux density, inversely as the 0.3 power.

The data quoted apply to receiving aerials of low directivity in which the main lobe of the radiation pattern covers a relatively large segment of the sky. With highly directional aerials beamed at a strong radio source, the noise intensity may increase considerably.

The increase in noise intensity may be evaluated as follows. Let a discrete radio source be visible from the site of the receiving

aerial at a solid angle Ω_{disc} and have an equivalent temperature T_e which exceeds the mean equivalent temperature of cosmic radio emission, T_{cosm}. If the polar diagram of the receiving aerial is described by a solid angle Ω, the equivalent temperature of the radio emission reaching the aerial may be found by the formula

$$T_e = T_{disc}\frac{\Omega_{disc}}{\Omega} + T_{cosm}\frac{\Omega - \Omega_{disc}}{\Omega} \qquad (6.12)$$

The solid angle of an aerial of gain G may be found by the formula

$$\Omega = 11.5\eta/G \text{ steradians} \qquad (6.13)$$

where η is the effective area of the aerial, usually equal to 0.6 to 0.8.

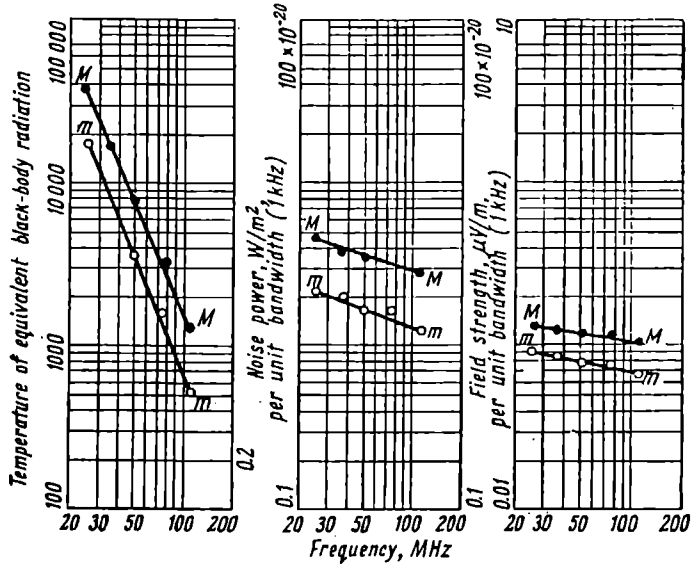


Fig. 6.8. Dependence of the maximum and minimum noise intensities on frequency. *M*, daily maximum; *m*, daily minimum

If the solid angle of the aerial is many times the solid angle Ω_{disc} of the discrete radio source, the first term in (6.12) may be neglected, and

$$T_e \cong T_{cosm}\,°\text{K} \qquad (6.12a)$$

Conversely, if $\Omega = \Omega_{disc}$, the second term reduces to zero, and

$$T_e \cong T_{disc}\,°\text{K} \qquad (6.12b)$$

that is, exceeds the mean temperature of cosmic radio emission. Naturally, when $\Omega < \Omega_{disc}$, the radiation reaching the aerial will have the same temperature T_{disc}.

Such a discrete source of strong radio emission is the sun. The emission of the sun may be of two types, quiet and disturbed. The dependence of the equivalent noise temperature of the quiet sun on frequency is represented by the upper curve in Fig. 6.9 [104]. The lower curve in the same graph represents a similar relationship

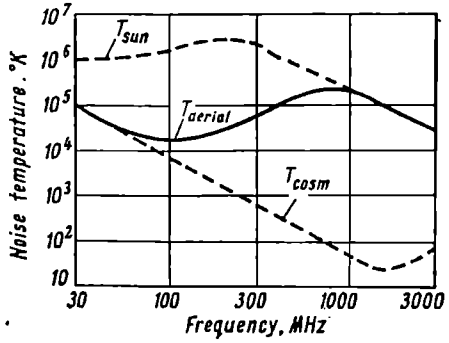


Fig. 6.9. Equivalent noise temperature of the quiet sun (the upper curve) and cosmic radio emission (the lower curve) as a function of frequency

for cosmic radio emission. As is seen, the equivalent noise temperature of the sun is markedly higher than that of cosmic radio-emission. If the sun is within the polar diagram of the aerial, the equivalent noise temperature of the sun may be found by Eq. (6.12) where the solid angle Ω_{disc} is given by

$$\Omega_{disc} = \pi\alpha^2/4 \text{ steradians} \qquad (6.14)$$

where α is the plane angle at which the discrete source is seen. For the sun, the angle α is slightly more than 30′ of arc because radio emission comes mainly from the sun's crown.

As to radio emission from the unquiet sun, it may come from sunspots. As sunspots pass through the central meridian, spikes appear in a record of solar radio emission, which may be hundreds of times stronger than the normal radiation. As a rule, spikes last a few seconds. Then there are "splashes" traceable to solar flares, the cause of SID events mentioned in § 5.13. Splashes usually occupy a few minutes, and the flux of radio emission may then be millions of times greater than the radio emission of the quiet sun.

6.4. Thermal Noise Due to the Atmosphere and Earth's Surface

This category of noise includes the thermal radiation produced by uncondensed water vapour and molecular oxygen in the lower atmosphere. From molecular physics it is known that if a gas selectively absorbs a radiation (resonance absorption) it will emit radiation at the same frequencies when hot. In the centimetric waveband two regions of resonance absorption are known, one at the 1.35-cm line of water vapour ($f = 22$ gigahertz) and the other at the 0.5-cm

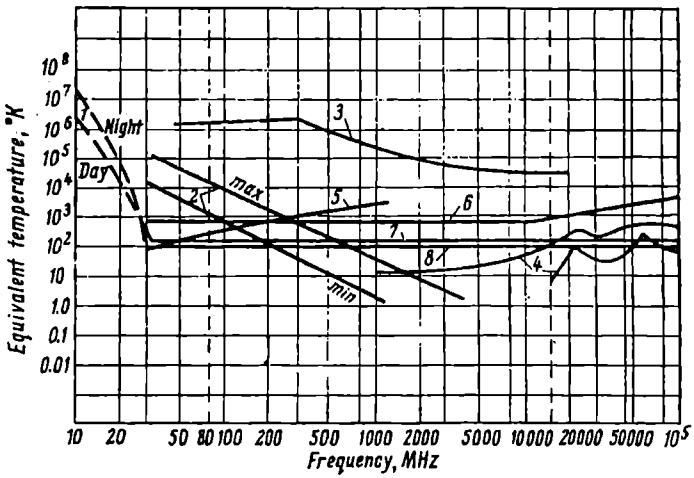


Fig. 6.10. Equivalent temperature of external and internal noise as a function of frequency

line of molecular oxygen ($f = 60$ gigahertz). The heated atmosphere emits these frequencies which are picked up by receivers as noise. As with atmospheric and cosmic noise, it is convenient to express the strength of this emission in terms of equivalent noise temperature.

The magnitude of noise produced by uncondensed water vapour and molecular oxygen can be found from the plot of Fig. 6.10 [105] relating the equivalent temperature of internal and external noise to frequency. The broken lines, *1*, represent the intensity of atmospheric noise in the range 10-30 megahertz by day and at night. The full lines, *2*, represent the daily maxima and minima of cosmic noise. Curve *3* represents the noise produced by the "quiet" sun. Two curves, *4*, describe the thermal noise produced by the atmosphere; the lower curve applies to a vertically directed aerial, and the upper curve to an aerial arranged at a small angle to the hori-

zontal. In the latter case the noise intensity is higher because a greater part of the atmosphere contributes to the overall noise. Both curves show distinct peaks at frequencies of 22 and 60 gigahertz.

From Fig. 6.10 it is seen that among external noise at frequencies upwards of two gigahertz, the thermal noise of the atmosphere dominates over cosmic noise.

To give an idea about the relative magnitudes of internal and external noise, the graph of Fig. 6.10 includes curves representing internal (receiver) noise. Curve *5* applies to present-day low-noise valves, curve *6* to present-day crystal mixers, curve *7* to future cooled crystal mixers, and curve *8* to masers and parametric amplifiers.

A further source of external noise is the earth's surface heated to an average of 300°K. With a horizontally directed aerial, the thermal radio emission of the earth is picked up to contribute to overall noise. Moreover, even when an aerial is directed vertically, the thermal emission of the earth affects the inevitable side lobes of the aerial's polar diagram. This is taken care of by curve *8* in Fig. 6.10.

To sum up, we may say that in the range from 30 to 200 megahertz, receiver sensitivity is limited by cosmic noise while in the range from 200 megahertz to 100 gigahertz the limiting factor is the noise of crystal mixers (radio emission from the quiet sun is neglected).

The situation is different with masers. From 30 to about 500 megahertz the limiting factor is cosmic noise; from 500 megahertz to 18 gigahertz this is the internal noise of masers and the thermal emission of the earth, and from 18 to 100 gigahertz, the thermal emission of the atmosphere.

Chapter Seven

Space Communication

7.1. Basic Definitions

The term 'space communication' applies to radio traffic between a ground station and a satellite or space probe, between satellites or space probes, and also between ground stations via a natural or man-made space body (the Moon, the Sun, Venus, or a communication satellite).

At the present-day state of the art, the interest lies mainly in radio communication with spacecraft and in the possibility of extending broad-band microwave (USW) line-of-sight transmission between remote points. Accordingly, this Chapter will be concerned only with these aspects of space communication.

In both cases, use is made of direct waves which are free to pass through the ionosphere and are propagated in practically straight-line paths.

For a satellite in a circular orbit around the earth there is a simple equation relating its height above the earth's surface and its period of orbit. This relation stems from the fact that at any instant the pull of gravity

$$F = kMm/r^2 \text{ N} \tag{7.1}$$

must be balanced by the centrifugal force

$$F_c = mv^2/r \text{ N} \tag{7.2}$$

where $k = 6.67 \times 10^{-11}$ m^3/kg-sec^2 is the gravity constant;
$M = 5.98 \times 10^{24}$ kg is the mass of the earth;
m = mass of the satellite;
r = radius of the circular orbit;
v = orbital velocity of the satellite.

Equating the expressions for F and F_c gives the linear velocity

of the satellite in a circular orbit

$$v = \sqrt{kM/r} \text{ m/s} \tag{7.3}$$

which, as is seen, does not depend on the mass of the satellite.

The orbital period of a satellite is given by

$$T = 2\pi r/v = 2\pi \sqrt{r^3/kM} \text{ s} \tag{7.4}$$

which is a mathematical expression for the Kepler's third law stating that the squares of the periods of revolution of the planets about the sun are proportional to the cubes of their mean distances from the sun.

Substituting the values of k and M and expressing T in minutes, the relation between the radius of a circular orbit and the orbital period is

$$T = r_{km}^{3/2}/6.02 \times 10^3 = 1.66 \times 10^{-4} r_{km}^{3/2} \text{ min} \tag{7.4a}$$

The height of the orbit, H, is given by

$$H = (r_{km} - 6370) \text{ km} \tag{7.5}$$

Among other things, the equations for T and H show that a "two-hour" satellite should be launched to a height of about 2000 kilometres, while a 24-hour satellite should be placed in an equatorial orbit with a height of 35,810 kilometres, called a *synchronous orbit* because it will appear stationary relative to a point of the earth if the satellite is travelling in the same direction as the rotation of the earth.

The equation for T has been derived on the assumption that the satellite travels in a vacuum. The atmosphere produces a drag and modifies the results.

For a satellite in an elliptical orbit, the distance r is a half-sum of the distances to the satellite at apogee and perigee. This distance is also equal to the major axis of the elliptical orbit.

Satellite observations have shown that the drag is strongest in the denser atmosphere lying at altitudes under 1500 kilometres. At greater altitudes, satellites experience practically no drag and can remain in orbit for years. A satellite placed in an elliptical orbit is subject to drag at perigee. As a result, its apogee is reduced, and the orbit approaches a circular one. In accordance with Kepler's third law, the period of revolution is then also decreased. On re-entering the denser atmosphere, satellites are heated and burn.

The orbital periods for satellites in circular orbits are summarized in Table 7.1.

For the purpose of radio communications, a satellite should stay in orbit for decades. Accordingly, it must be launched to an altitude in excess of 1500 kilometres. Another essential point is

Table 7.1. PERIODS OF ORBIT FOR SATELLITES IN ELLIPTICAL ORBITS

Height above earth, km	Linear velocity, m/sec	Period of orbit, min	Height above earth, km	Linear velocity, m/sec	Period of orbit, min
200	7791	88	3000	6525	151
300	7732	90	4000	6203	175
400	7675	92	5000	5924	201
500	7619	94	6000	5679	228
1000	7356	105	7000	5473	257
2000	6903	127			

that the communication satellite should be simultaneously visible from various points on the earth's surface. Therefore, the orbit of a satellite should be properly oriented relative to the earth. A satellite may thus be placed in a circular polar, a circular equatorial, or an elliptic inclined orbit.

A satellite in a circular polar orbit will travel over both poles during each circuit. If the orbital period is not a whole sub-multiple of twenty-four hours, the projection of the satellite's trajectory on the earth's surface will continually move, and sooner or later the satellite will travel over any specified point on the ground. Conversely, a satellite in a circular equatorial orbit will stay over the equator and will be visible within a strip bounded by certain northern and southern latitudes. A satellite in an inclined orbit, making an angle ψ with the equator, will pass over points at latitudes $+\psi$ and $-\psi$.

7.2. General Principles of Satellite Communications

Satellites can be used as active and passive retransmitters in long-distance communications systems.

An active communication satellite is one carrying an aerial system, a receiver, a transmitter, and power supplies. Some of the active communication satellites carry message-recording, play-back and programming facilities. A passive communication satellite is a metallic or metallized sphere reflecting radio signals back to earth well.

The most crucial question in the use of a satellite as a retransmitter in a long-distance communication system is the choice of operating frequency. Obviously, this must be such that the signal will be free to pass through the ionosphere, on the one hand, and will not be attenuated in the troposphere, on the other.

The lower limit for operating frequency is the MUF for ground-based stations. In other words, use must be made of frequencies

exceeding the MUF. The average figure is 10 megahertz, although it may be as high as 80 or even 100 megahertz when the sun is active, and as low as two megahertz for the polar regions during the polar

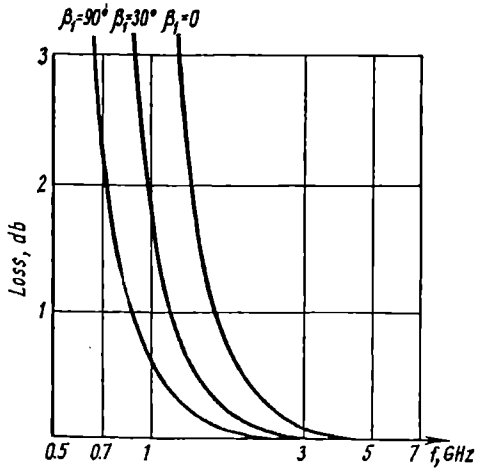


Fig. 7.1. Calculated one-way Faraday loss for a signal passing through a homogeneous ionized region, three angles of elevation and as a function of frequency

night. Even the frequencies upwards of the critical one suffer absorptive losses as they pass through the *D* and *E* regions. However, as is seen from Table 7.2, the absorptive losses are markedly decreased with an increase of frequency.

Table 7.2. RADIO ABSORPTION IN THE *D* AND *E* REGIONS

Frequency, MHz	30	50	100
Loss, db	2.5	1.0	0.3

Besides, the ionosphere causes rotation of the plane of polarization (Faraday rotation). For an aerial capable of receiving a plane-polarized wave, this rotation amounts to a loss of energy (Faraday loss). The situation is aggravated by the fact that the plane of polarization continually varies its orientation due to incessant fluctuation in the electron density.

Fig. 7.1 gives the calculated one-way Faraday loss [105] as the signal passes through a homogeneous ionized region. The data have

been calculated for three angles of elevation and as a function of frequency. It has been assumed that $N = 2.8 \times 10^{12}$ per cubic metre in the interval from 230 to 370 kilometres and that $N = 0$ at any other height. The strength of the longitudinal magnetic field has been set at $H_l = 30$ amperes per metre.

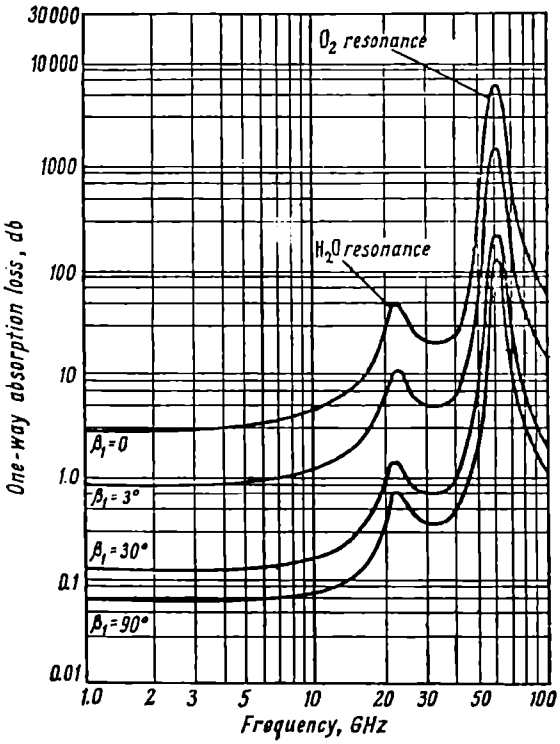


Fig. 7.2. Loss due to molecular absorption by atmospheric oxygen and uncondensed water as a function of frequency

An obvious counter-measure against Faraday rotation is to use transmitting and receiving aerials with circular polarization. From Fig. 7.1 it is seen that circular polarization is a "must" on frequencies up to 1.42 gigahertz. In the interval from 1.4 to 7 gigahertz, the loss can be found from the plot of Fig. 7.1. At frequencies in excess of 7 gigahertz Faraday loss may be neglected.

The upper limit of frequencies suitable for long-distance communication via satellites is decided by the absorption by precipitation particles and is, as was explained in § 3.14, about 6 gigahertz. If ground-based stations are in localities with low precipitation rates,

the upper limit is decided by the absorption by uncondensed water vapour and molecular oxygen of the atmosphere.

The loss in the troposphere as a function of frequency is shown in the graph of Fig. 7.2 plotted for four angles of elevation. It is seen that the limiting frequency is about 10 megahertz.

To sum up, long-distance communications systems via satellites should preferably use frequencies from 3 to 6 gigahertz. As follows from Fig. 6.10, the main source of interference in this frequency range is internal (receiver) noise and the thermal radiation of the earth, and this may be successfully controlled by the use of masers and aerials with effective sidelobe attenuation.

7.3. Energy Considerations in Satellite Communications Systems

The simplest type of active communication satellite is one which records messages as it travels over a ground-based transmitting station using a magnetic tape recorder or some other equipment and re-transmits them in passing over the receiving station on radio command from the ground or from a programming device.

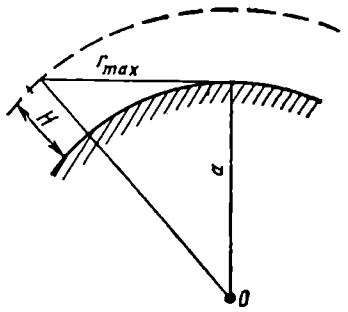


Fig. 7.3. Maximum distance of an earth satellite in a circular orbit at a height H

This method is simplest in that it reduces to operation within the line of sight over relatively short distances. For a satellite placed at a height of 5000 kilometres, the longest distance (Fig. 7.3) from the earth will be

$$r_{max} = \sqrt{(a+H)^2 - a^2} = 9400 \text{ km}$$

A very rough estimate shows that at a frequency of 3 gigahertz ($\lambda = 10$ centimetres), which is the lower limit of the optimum band, the basic transmission loss, Eq. (1.21a) is,

$$L_{0,f} = 180 \text{ db}$$

To receive signals with a bandwidth of 5 megahertz and a signal-to-noise ratio of 20 decibels, the power at the input to the receiver must be

$$P_2 = -115 \text{ dbW}$$

Assuming that the communication satellite uses an isotropic aerial with a gain of $G_1 = 1$ and that the aerial on the ground has

a gain of $G_2 = 10^5$ ($G_2 = 50$ decibels), and re-writing Eq. (1.55a) as

$$P_{1,\,\mathrm{dbW}} + G_{1,\,\mathrm{db}} + G_{2,\,\mathrm{db}} = P_{2,\,\mathrm{dbW}} + L_{0,\,f,\,\mathrm{db}} \qquad (7.6)$$

and because the attenuation function $F = 0$ decibels, we get

$$P_1 = -115 + 180 - 50 = 15 \text{ dbW}$$

which works out to a transmitter power of 30 watts. That is, use may be made of a low-power transmitter and very simple aerials.

Another type of active communication satellite is one carrying the equipment of an ordinary microwave-link repeater station and intended for direct real-time relay of signals. Such a satellite must be within the range of mutual visibility of the transmitting and receiving stations. Accordingly, the calculations performed fully apply to this case, too.

There is a good deal of attraction in using synchronous (24-hours) satellites for communications. Since such a satellite appears stationary, there is no need to use ground-based satellite-tracking aerials. The transmitter power for a stationary satellite is determined in the example that follows.

Example 7.1. Determine the power of a transmitter for an active communications satellite placed in the 24-hour orbit. The system is to use a wavelength of 10 cm. The satellite is to use a directional aerial of $G_1 = 10^3 = 30$ db. The ground-based aerial will have a gain of $G_2 = 10^4 = 40$ db. The power at the input to the receiver must be $P = -95$ dbW for a signal bandwidth of 5 MHz, a signal-to-noise ratio of 40 db, and a receiver noise figure $N = 3$ db.

Solution. A stationary satellite will be in a 36,000-km orbit. From Eq. (1.21a) we find that the basic transmission loss in free space is $L_{0,\,f} = 193$ db.

Substituting the numerical values of P_2, $L_{0,\,f}$, G_1, and G_2 in (1.55a) and setting $F_{\mathrm{db}} = 0$, we get

$$P_1 = -95 + 193 - 30 - 40 \quad 28 \text{ dbW}$$

which corresponds to a transmitter power of about 600 watts.

The example bears out the feasibility of a communications system using a synchronous satellite.

The problem is more involved with a passive communication satellite which is a metallized sphere in the simplest case. It reflects the signals coming from a ground-based station back to earth, where they are picked up by receiving aerials at another point. As will be shown shortly, such a communication satellite is less efficient in terms of energy consumption, so that larger transmitters and automatically tracking aerials must be used for reliable operation.

A metallic sphere of diameter d has, to a first approximation, an effective scattering cross-section per unit volume given by

$$\sigma = \pi d^2/4 \text{ m}^2 \tag{7.7}$$

provided $d \gg \lambda$, which is true in our case.

Writing r_1 and r_2 for the distances of the satellite from the transmitter and receiver respectively and using the usual notation, the power flux density at the satellite will be given by

$$S_{sat} = P_1 G_1/4\pi r_1^2 \text{ W/m}^2 \tag{7.8}$$

Under the circumstances, the equivalent power of the re-radiated emission from the satellite will be

$$P_{sat} = S_{sat}\sigma \text{ W} \tag{7.9}$$

The power flux density due to the re-radiated emission at the receiving aerial will be

$$S_2 = P_{sat}/4\pi r_2^2 \text{ W/m}^2 \tag{7.10}$$

Since the effective area of the receiving aerial is

$$A_2 = G_2\lambda^2/4\pi \text{ m}^2 \tag{7.11}$$

the power at the input to the receiver will be given by

$$P_2 = S_2 A_2 \text{ W} \tag{7.12}$$

Combining (7.8) through (7.12) gives the following expression for the total transmission loss

$$L = P_1/P_2 = 64\pi^3 r_1^2 r_2^2/G_1 G_2 \lambda^2 \sigma \tag{7.13}$$

Substituting the expression for σ in (7.13) gives

$$L = (16\pi r_1 r_2/\lambda d)^2 (1/G_1 G_2) \tag{7.14}$$

or in decibels

$$L_{\text{db}} = 20 \log_{10} \left(\frac{16\pi r_1 r_2}{\lambda d}\right) - G_{1,\text{db}} - G_{2,\text{db}} \tag{7.14a}$$

The basic path loss is

$$L_{0,\text{db}} = 20 \log_{10} (16\pi r_1 r_2/\lambda d) \tag{7.15}$$

Example 7.2. Determine the power of a ground-based transmitter for use a communications system built around a passive repeater satellite, which is to operate on $\lambda = 6$ cm. The satellite is to be placed in a circular polar orbit at 3000 km above the earth. With this orbit and a propagation-path length of 4000 km the worst-case distances are $r_1 = r_2 = 6140$ km. $G_1 = G_2 = 60$ db. Use is made

of maser amplifiers with an equivalent noise temperature of 150°K and a signal bandwidth of 5 MHz. With a signal-to-noise ratio of 5 db, the input power is $P_2 = -135$ dbW.

Solution. Substituting the specified numerical values in (7.15) gives

$$L_0 = 304 \text{ db}$$

Using Eq. (1.21a), we get

$$P_1 = -135 + 304 - 60 - 60 = 49 \text{ dbW}$$

which corresponds to a transmitter power of about 81 kilowatts.

7.4. Prospects of World-wide Satellite Communications Systems

A simple geometric reasoning will show that three active synchronous satellites placed in an equatorial orbit and spaced an equal distance apart may form a world-wide communications system—the one reaching any point on the earth's surface. However, the information-handling capability of such a system will be limited because there must be a separate transmitter for each channel.

Another category of world-wide satellite communications systems is based on passive-repeater satellites visible from any point on the earth, which can be used for retransmission of an unlimited number of programmes at a time.

The high cost of launching of a metallized sphere serving as a passive reflector and the short time during which such a sphere is simultaneously visible from the terminal points of a radio circuit make the use of a single communication satellite for the purpose unattractive. Instead, there must be a system of passive satellites giving global coverage. Their number and trajectories should be chosen so that there will be at least one satellite over a given radio circuit at any instant of time.

At least two approaches can be used to tackle the problem: (1) to launch a system of so-called positioned satellites and (2) to launch a system of random-orbit satellites.

In the former case, the satellites are to be launched in precise, controlled equatorial orbits at precisely equal time intervals. The number and trajectories of satellites must be such that they will girdle the globe or, at least, the area where radio communications are to be provided. Obviously, this is a formidable task both technically and economically. Besides, atmospheric drag would inevitably distort the orbits, and operation of the system would be disorganized with time.

The second approach requires a large number of passive reflectors, but it is simpler to carry out.

As set forth by Pierce and Kompfner [106], [107], the underlying idea is as follows. A number of satellites, n, are launched at random time intervals to uniform-orbit altitudes. The possibility of radio communications is based on the fact that with a sufficiently large number of satellites there will always be the probability of at least one satellite being seen over a given radio link.

The number of satellites, n, required for reliable communication along a given radio link is simple to determine. Let p denote the probability that a satellite is visible from both terminals of a given radio link. Then $(1 - p)$ will be the probability of the satellite not being seen. Hence, the probability of not seeing any of n satellites will be

$$S = (1 - p)^n \tag{7.16}$$

where S is, in effect, the probability of communication outage, and $(1 - S) \times 100$ per cent is the communication reliability.

Taking the logarithms of both sides of (7.16), the required number of satellites is obtained

$$n = \frac{\log_{10} S}{\log_{10} (1 - p)} \tag{7.17}$$

Example 7.3. Determine the number of random-orbit satellites required to ensure radio communications during 90, 99, and 99.9 per cent of the time on the assumption that the probability of a satellite being seen from both terminals of the link is $p = 0.055$.

Solution. The respective values of S are 0.10, 0.01 and 0.001. Substituting them and the value of p in (7.17) gives $n = 40$ satellites to maintain radio communications during 90 per cent, 80 satellites for 99 per cent, and 120 satellites for 99.9 per cent of the time.

7.5. Ground-to-Space and Space-to-Space Communications

For radio communications between ground-based stations and satellites the optimum traffic frequencies are found from the same considerations as set forth in § 7.2. The required transmitter power (ground-based or satellite-borne) can be determined by means of (7.6).

Example 7.4. Determine the required power for a satellite-borne transmitter to ensure radio communications at distances of 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, 10^9 and 10^{10} km at a frequency of 3 GHz using a transmitting aerial with a gain of 20 db and a receiving aerial with a gain of 60 db. The power at the receiver input should

be −173 dbW for a signal bandwidth of 100 Hz, $T_e = 100°K$, and a signal-to-noise ratio of 6 db.

Solution. The results are summarized in Table 7.3.

Table 7.3

r, km	10^3	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}
P_1, dbW	−70	−50	−30	−10	+10	+30	+50	+70

For radio communications between two spacecraft outside the terrestrial atmosphere the limitations imposed on the choice of frequencies by the transparency of the ionosphere and tropospheric absorption no longer hold, and any frequency may in principle be used. The most attractive choice is millimetric waves and optical frequencies owing to the compact size of the equipment, especially aerials, involved.

REFERENCES

1. Ohman, G. P. *IRE Transactions on Antennas and Propagation*, vol. AP-5, Jan., 1957, pp. 140-142.
2. Ohman, G. P. *IRE Transactions on Antennas and Propagation*, vol. AP-10, July, 1962, pp. 450-452.
3. Kashprovsky, V. E. *Geomagnetizm i aeronomiya*. vol. 3, 1963, No. 2, pp. 297-308 (distribution of local soil conductivities in the USSR).
4. Vvedensky, B. A. *Vestnik teoreticheskoi i eksperimental'noi elektrotekhniki*, 1928, No. 12, pp. 439-446 (propagation of ultra-short waves).
5. Vvedensky, B. A., Astafiev, A. V., Arenberg, A. G. *Ibid.*, 1928, No. 12, pp. 447-451 (radio communication with ultra-short waves).
6. Bullington, K. *Proc. IRE*. vol. 35, 1947, pp. 1122-1136.
7. Leontovich, M. A. In *Issledovaniya po rasprostraneniyu radiovoln* (Studies into Radio Propagation), collected volume No. 2, USSR Academy Press, 1948, pp. 5-12.
8. Shuleikin, M. V. *Rasprostranenie elektromagnitnoi energii* (Propagation of Electromagnetic Energy), First Russian Radio Bureau Press, 1923.
9. Burrows, C. R., Gray, M. C. *Proc. IRE*, vol. 29, 1941, pp. 16-24.
10. Shchukin, A. N. *Rasprostranenie radiovoln* (Radio Wave Propagation). Moscow, Svyazizdat, 1940.
11. Eckersley, T. L. *Proc. IRE*, vol. 18, 1930, p. 1160.
12. Dolukhanov, M. P. *Elektrosvyaz*, 1939, No. 4 (field strength calculation for surface waves propagated over multi-section paths).
13. Millington, G. *Journal of IEE*, vol. 96, part III, 1949, p. 53.
14. Grinberg, G. A. *Journal of Phys.*, vol. 6, 1942, p. 185 (on the coast-line refraction of radio waves).
15. Grinberg, G. A., Fock, V. A. In *Issledovaniya po rasprostranenityu radiovoln* (Studies into Radio Propagation), collected volume No. 2, USSR Academy Press, 1948, pp. 69-96.
16. Feinberg, E. L. *Izvestiya Akademii Nauk SSSR*, vol. 7, 1944, p. 167.
17. Feinberg, E. L. *Radiotekhnika*, 1950, No. 4, pp. 3-16.
18. Zenneck, J. *Ann. d. Phys.*, vol. 23, 1907, p. 846.
19. Norton, K. A. *Proc. IRE*, vol. 25, 1937, pp. 1192-1202.
20. Ryazin, P. A. In *Novetshie issledovaniya rasprostraneniya radiovoln* (Latest Studies into Radio Wave Propagation), collected volume No. 1, GTTI, 1945, p. 101.
21. Mandelstamm, L. I., Papaleksi, E. Ya., Shchegolov, E. Ya. *Ibid.*, p. 145.
22. Smith-Rose, R. L. *Proc. IRE*, vol. 38. 1950, pp. 16-20.
23. *Bolshaya Sovetskaya Entsiklopediya*, 2nd ed., vol. 45, p. 71 (physical constants).
24. Domb, C., Pryce, M. H. *Journ. of IEE*, vol. 94, part III, 1947, pp. 325-336.
25. Watson, G. N. *Proc. Roy. Soc. A*, vol. 95, 1918, pp. 83-99.
26. Vvedensky, B. A. *Zhurn. teor. i eksper. fiziki*, vol. 6, 1936, p. 168; vol. 6, 1936, p. 1836; vol. 7, 1937, p. 1647 (on the diffraction propagation of radio waves).

27. Bremmer, H. *Terrestrial Radio Waves.* Elsevier Publ., Amsterdam, 1948.
28. Fock, V. A. *Difraktsiya radiovoln vokrug zemnoi poverkhnosti* (Diffraction of Radio Waves around the Earth's Surface). USSR Academy Press, 1946.
29. Belkina, M. G. *Tablitsa dlya vychisleniya elektromagnitnogo polya v oblasti tent dlya razlichnykh pochv* (Table of Electromagnetic Fields in the Shadow Region for Various Soils). "Sovetskoe Radio", Moscow, 1949.
30. Leontovich, M. A., Fock, V. A. *Zhurn. teor. i eksper. fiziki*, vol. 16, 1946, No. 7, p. 557.
31. Fock, V. A. *Ibid.*, vol. 19, 1949, No. 10, p. 916.
32. *Tables of Modified Hankel Functions of Order One-Third and Their Derivative.* Ann. Harvard Univ. Computation Laboratory, vol. 2, 1945.
33. *Transactions of the 8th Plenary Assembly of the CCIR in Warsaw*, 1956.
34. *Rasprostranenie ultrakorotkikh voln* (Propagation of Ultra-Short Waves). Translations from English, ed. by B. A. Shillerov, "Sovetskoe Radio", Moscow, 1954.
35. Feinberg, E. L. *Rasprostranenie radiovoln vdol' zemnoi poverkhnosti* (Propagation of Radio Waves along the Earth's Surface). USSR Academy Press, 1961.
36. Landsberg, G. S. *Optika* (Optics). 3rd ed., 1954.
37. Schelling, J., Burrows, C., Ferrel, E. *Proc. IRE*, vol. 21, 1933, pp. 426-463.
38. Dickson, F. H., Egli, J. J., Herbsreit, J. W., Wickizer, G. S. *Proc. IRE*, vol. 41, 1953, pp. 967-969.
39. Vvedensky, B. A., Ponomarev, M. I. *Izvestiya Akademii Nauk SSSR (engineering division)*, 1946, No. 9, p. 1201.
40. Fock, V. A. In *Issledovaniya po rasprostraneniyu radiovoln* (Studies into Radio Wave Propagation), collected volume No. 2. USSR Academy Press, 1948, p. 40.
41. Krasnushkin, P. E. *Metod normalnykh radiovoln v primenenii k probleme dalnei radiosvyazi* (The Normal-Mode Solution as Applied to Long-Distance Radio Communications). Moscow University Press, 1947.
42. Booker, H. G. Walkinshaw, W. *Phys. Soc. Special Report on Meteorological Factors in the Radio-Wave Propagation*, 1946, pp. 80-127.
43. Pekeris, C. L., Ament, W. S. *Phil. Mag.*, vol. 38, 1947, pp. 801-823.
44. Brekhovskikh, L. M. *Izvestiya Akademii Nauk SSSR* (physics series), vol. 13, 1949, p. 409.
45. Brekhovskikh, L. M. *Volny v sloistykh sredakh* (Waves in Stratified Media). USSR Academy Press, 1957.
46. Krasilnikov, V. A. *Izvestiya Akademii Nauk SSSR* (geography and geophysics series), vol. 13, 1949, No. 1, p. 33.
47. Dolukhanov, M. P. *Radiotekhnika*, 1958, No. 1, pp. 49-63.
48. Villars, F., Weisskopf, V. F. *Phys. Rev.*, vol. 94, 1954, p. 232.
49. Villars, F., Weisskopf, V. F. In *Voprosy dalnei svyazi na UKV* (Problems of Long-Distance USW Radio Communications), collected volume. "Sovetskoe Radio", Moscow, 1957.
50. Troitsky, V. N. *Radiotekhnika*, 1956, No. 5, p. 3.
51. Friis, H. T., Crawford, A. B., Hogg, D. C. *Bell Syst. Techn. Journ.*, vol. 36, 1957, pp. 627-645.
52. Norton, Vogler, Mansfield, and Short. In *Voprosy dalnei svyazi na UKV* (Problems of Long-Distance USW Radio Communications), collected volume "Sovetskoe Radio", Moscow, 1957, pp. 16-183.
53. *Fizichesky entsiklopedichesky slovar'* (Encyclopedic Dictionary of Physics). Moscow, 1962, vol. 2, pp. 35-36.
54. Bean. Russian translation in *Voprosy radiolokatsionnoi tekhniki*, 1956, No. 5.
55. Yeh, L. P. *British Communications and Electronics.* Nov., 1958, pp. 707-716.

56. Gaertner, H. Transparency of a Turbid Atmosphere to Infra-Red Waves. Moscow, 1949.
57. Tolbert, C. W., Straiton, A. W. *Proc. IRE*, vol. 49, No. 3, March, 1961.
58. Miller, S. E. *Proc. IRE*, vol. 50, 1962, p. 1216.
59. Luck, D. G. *RCA Rev.*, Sept., 1961, pp. 359-409.
60. Istomin, V. G. In *Iskusstvennye sputniki zemli* (Earth Satellites), collected volume, issue 7, 1961, pp. 64-67.
61. *Ibid.*, issue 4, 1960, pp. 171-183.
62. *Ibid.*, issue 11, 1961, pp. 94-97.
63. *Fizichesky entsiklopedichesky slovar'* (Encyclopedic Dictionary of Physics). Moscow, 1962, vol. 1, pp. 98-100.
64. Krasovsky, V. I., *et al.* In *Iskusstvennye sputniki zemli* (Earth Satellites), collected volume, issue 6, 1961, pp. 113-126.
65. Krasovsky, V. I. In *Iskusstvennye sputniki zemli* (Earth Satellites), collected volume, issue 2, 1958, p. 36.
66. Kryuchkov, S. I. *Zhurn. prikl. fiziki*, vol. 7, issue 3, 1930, p. 61.
67. Chapman, S. *Proc. Phys. Soc.*, vol. 43, 1930, p. 26.
68. Ginzburg, V. L. *Rasprostranenie elektromagnitnykh voln v plazme* (Propagation of Electromagnetic Waves in the Plasma). Fizmatgiz, Moscow, 1960.
69. Gringauz, K. I., Rudakov, V. A. In *Iskusstvennye sputniki zemli* (Earth Satellites), collected volume, issue 6, 1961, pp. 48-62.
70. *Handbook of Geophysics*. New York, 1960.
71. *Ionospheric Sporadic E*. Ed. by E. K. Smith, S. M. Matsushita. Pergamon Press, Oxford, 1962.
72. Vernov, S. N., *et al. Doklady Akademii Nauk SSSR*, vol. 125, 1959, p. 304.
73. Van Allen, J. A., Frank, L. A. *Nature*, vol. 183, 1959, p. 430.
74. Vernov, S. N., *et al.* In *Iskusstvennye sputniki zemli* (Earth Satellites), issue 5, 1960, pp. 24-29.
75. Gringauz, K. I., *et al. Ibid.*, issue 6, 1961, pp. 101-107.
76. Gringauz, K. I., *et al. Ibid.*, pp. 108-112.
77. Kabanov, N. I. *Radiotekhnika i elektronika*, vol. 5, 1960, No. 10, pp. 1576-1592.
78. *Mesyachnye prognozy rasprostraneniya radiovoln na 1962* (Monthly Predictions of Radio Wave Propagation for 1962). USSR Academy of Sciences, 1961-62.
79. Alfvén, H. *Cosmical Electrodynamics*. Oxford, 1950.
80. Helliwell, R., Morgan, M. G. *Proc. IRE*, vol. 47, 1959, pp. 200-208.
81. Storey, L. R. O. *Phil. Trans. Roy. Soc., A*, vol. 246, 1953, pp. 113-141.
82. Alpert, Ya. L. *O rasprostranenii elektromagnitnykh voln nizkoi chastoty nad zemnoi poverkhnostyu* (On the Propagation of Low-Frequency Waves along the Earth's Surface). USSR Academy of Sciences, 1955.
83. Alpert, Ya. L. *Rasprostranenie radiovoln i ionosfera* (Radio Propagation and the Ionosphere). USSR Academy of Sciences, 1960.
84. Wait, J. R. *Proc. IRE*, vol. 45, 1957, pp. 760-767.
85. *Rev. UER, A*, No. 73, 1962, pp. 109-121.
86. Hess, H. A. *Proc. IRE*, vol. 40, 1952, pp. 1065-1068.
87. Bartenev, G. M. *Izvestiya Akademii Nauk SSSR* (engineering division), No. 9, 1947, pp. 1153-1171.
88. Bartenev, G. M. *Doklady vysshei shkoly*, No. 4, 1958.
89. Shuleikin, M. V. *Izvestiya Akademii Nauk SSSR* (engineering division), 1938, No. 5.
90. Kazantsev, A. N. *Izvestiya Akademii Nauk SSSR* (engineering division), 1946, No. 9, pp. 1261-1296.
91. Kazantsev, A. N. *Ibid.*, 1947, No. 9, pp. 1107-1137.
92. Kazantsev, A. N. *Paper at the 8th Plenary Assembly of CCIR*, Warsaw, 1956.

93. *Report of a Working Group, CCIR.* Document VI/21, Warsaw, 1961.
94. Bailey, D. K., *et al. Phys. Rev.*, vol. 86, 1952, pp. 141-145.
95. *Proc. IRE*, vol. 48, 1960, pp. 4-44 (radio transmission by ionospheric and tropospheric scatter).
96. Dolukhanov, M. P. *Dal'nee rasprostranenie UKV* (Long-Distance USW Propagation), Moscow, 1962.
97. Forsyth, P. A., Vogan, E. L., *et al. Proc. IRE*, vol. 45, 1957, pp. 1642-1657.
98. Egorov, K. P. *Osnovy mnogokanal'not svyazi* (Principles of Multi-Channel Communications). Moscow, 1962.
99. *Zarubezhnaya radioelektronika*, 1961, No. 11, pp. 3-15.
100. *Ibid.*, pp. 16-26.
101. Stratton, J. A. *Electromagnetic Theory.* New York, 1941.
102. *Fizichesky entsiklopedichesky slovar'* (Encyclopedic Dictionary of Physics). Moscow, 1962, vol. 1, pp. 100-102.
103. Cottony, H. V. *Proc. IRE*, vol. 40, 1952, pp. 1053-1060.
104. Blattner, D. J. *IRE, Trans. Commun. Syst.*, CS-8, 1960, pp. 169-172.
105. Pratt, H. J. *IRE Trans. Commun. Syst.*, CS-8, 1960, pp. 214-221.
106. Pierce, J. R., Kompfner, R. *Proc. IRE*, vol. 47, 1959, pp. 372-380.
107. Pierce, J. R. *Bell Labor. Record*, vol. 37, 1959, pp. 323-329.

INDEX

Absorption, in auroral zone, 309
molecular, 201
polar-cap, 309
resonant, 201
selective, 201
Absorption coefficient, 27
extraordinary ray, 250
ordinary ray, 250
Advection, 156
Aerial, anti-fading, 287
directional, 22
directive gain of, 22
effective aperture of, 24
equivalent length, 25
inverted-L, 68
polar diagram of, 22
power gain of, 22
Aerial height, minimum effective, 58
relative, 103
Airy integral functions, 102
Angle, Brewster, 36
grazing, 32
polarizing, 32
pseudo-Brewster, 36
scattering, 165
Angstrom unit, 11
Antipode effect, 278
Atmosphere, completely mixed, 213
constant-constitution, 213
terrestrial, 211
Atmospheric refraction, 144, 195
forms of, 151
Atmospherics, 209, 337
Attachment coefficient, 227
Attenuation, by molecular scattering, 204
by precipitation particles, 196
by solids in atmosphere, 206
in imperfectly conducting medium, 26-30
in troposphere, 195
Attenuation function, 37
dependence on elevation angle, 53
four-ray propagation, 136
mixed-path, 82, 86
Auroral displays, 209
Austin-Cohen formula, 281

Barometric equation, 213
Beam waveguide, 334
Black-outs, radio, 281
Bragg equation for turbulent scattering, 170
Bragg law, 169
Brewster angle, 36
Burst communications systems, 329, 331

Calculation, short-wave radio circuit, 310
Kazantsev method, 315
Cauchy's criterion, 261
Centimetric waves, propagation of, 333
Chapman ionized layer, 225
Coastal refraction, 80, 90
nomogram of, 92-3
theory of, 91
Communication channel, requirements for, 40

Communication satellite, active, 353, 357
Communication, space, 351
Communications systems, burst, 329, 331
Complex dielectric constants, 20
Complex permittivity, 20
  relative, 20
Continuity equation, 226
Convection current, 21
Cosmic noise, 344
  radio, 345
Cosmic rays, 219
Critical frequency, 229, 254
Cross-section, scattering, 165
Current, convection, 21
  displacement, 21
Curvature, wave-path, 145

Decimetric waves, propagation of, 333
Design technique, simplified, 58
Dielectric constant, complex, 20
Diffraction, four-ray model of, 135
Dipole moment, permanent molecular, 142
Direct ray, 45, 58
Direction-finding errors, due to coastal refraction, 90
Dispersive medium, 239
Displacement current, 21
Distance, numerical, 74
  scale, 103
Diversity reception, 187
  frequency, 187
  space, 187
Doppler shift, 39
Duct, tropospheric, 16
Duct mode, 158
Ducting, 158
  tropospheric, 159
    ray-tracing approach to, 159

Earth, effects on radio propagation, 21
  effective radius of, 146
Earth surfaces, electrical characteristics of, 42
Echo, short-wave, 301
Eckersley's method, 81
Eddies, 161
Electron concentration, 226
Electron number density, 226
Equation of state, 212
Equivalent surface, 42
Ergodic theorem, 163
Es layer, 234
Excess refractive index, 143
  vertical profile of, 143
Extraordinary ray, 245, 247
  absorption coefficient, 250

Factor, effective earth's radius, 148
Fading, 39, 151, 161, 285
  control of, 151, 187
  depth of, 175
  distribution function of, 176
  Faraday, 297
  frequency selectivity of, 187
  interference, 295
  in tropospheric scatter, 174
  on short waves, 295
  rate of, 175
  Rayleigh distribution, 177
  spatial selectivity of, 186
  time selectivity of, 184
Faraday loss, one-way, 354
Faraday rotation, 354
Field, diffracted, 107
Field structure, 63
Feinberg's equations, 84
Flat-earth problem, 43, 60
Flow, laminar, 161
  turbulent, 161
Fog, basic characteristics of, 197
Forward scatter, 171
Free space, 13
Frequency, optimum-traffic, 311
  units, 11
Fresnel zone, 62, 119, 121-27
  clearance, 126-29

Geometrical optics, application to radio waves, 44
Goubau's beam waveguide, 334
Grazing angle, 32
Ground-reflected ray, 45
Ground wave, behaviour of, 87
  propagation of, 41
    effect of troposphere on, 150
    over flat earth, 43
  range of, 301
Group velocity, 238
Gyro-frequency, 243
  longitudinal, 244
  transverse, 244
Gyromagnetic resonance, 243

Hankel approximation, 102
Hankel function of order one-third, 102
Height, scale, 103, 213
Hertzian vector, 101
Huygens' principle, 61, 119, 187

Illuminated zone, 101
Image principle, 71
Index of refraction, excess, 142
Interference equations, 53
International standard atmosphere, 141
Ionization, by absorption of radiation, 215
  by collision, 215
  in homogeneous atmosphere, 220-25
  in real atmosphere, 228
  mechanisms of, 215
  peak, 221
  photo-, 215
Ionization frequency, 215
Ionization wavelength, 215
Ionized layer, Chapman, 225
  simple, 225
  virtual height of, 259
Ionizing agencies, 218
Ionosphere, 16-7
  constitution, 209
  reflection from, 17
Ionosphere storms, 281, 289
Ionospheric layers, 229, 267
  characteristics of, 231
Ionospheric maps, 267
Ionospheric propagation, 271
Ionospheric reflections, metric waves, 323
Ionospheric regions, 229, 267
  characteristics of, 231
  critical frequencies of, 232
  virtual heights of, 232
Ionospheric scatter, metric waves, 324
  signal-level variations in, 326
Ionospheric sounding, 209, 259
  oblique incidence, 209, 259
  oblique-incidence backscatter, 209, 259
  panoramic, 261
  vertical-incidence, 13, 209, 259
Ionospheric wave, 17
Isopleths of critical frequencies, 267
Isopleths of refractive index, 191
Isotropic radiator, 21

Kabanov effect, 262
Kepler's third law, 352

Lapse rate, standard, pressure, 141
  standard, temperature, 141
Laser, 206
Laser communication, 208
Layered irregularities, 173
Leontovich's approximate boundary conditions, 60
Lightning, leader, 338
  main discharge, 338
  pilot streamer, 338
  predischarge, 337
Line-of-sight propagation, 93
  over hills, 124

Line-of-sight radio-relay systems, 12
Line-of-sight range, 93
Lorentz force, 242
Loss, path, 24
transmission, 24
basic, 24
Lyman-alpha radiation, 220
Lyman-beta radiation, 220

Maps, surface-conductivity, 42
Maximum usable frequency, 254, 263
Maxwell's first equation for imperfectly conducting medium, 20
Mechanical wave number, 161
Medium, dispersive, 239
Meteor-scatter links, 332
Meteor trails, 220, 328
overdense, 329
underdense, 329
Metric waves, modes of propagation for, 332
MHD waves, 9, 271
Microwaves, 11
Microwave link, in hilly locality, 126
Mie's scattering, 200
Millimetric waves, propagation of, 333
Millington's method, 83
Minimum discernible power, 38
Minimum effective aerial height, 58
Mixed-path problem, 84
Mixing-in-gradient theory, 171
Moisture pockets, 155
MUF daily schedule, 313
MUF maps, 311
Multi-pathing, 174, 295
Multi-path interference, 278

Noise, atmospheric, 337
cosmic, 337, 344
radio, 345
galactic, 345
thermal, 349
Non-central Rayleigh density function, 184
Normal-mode method, 159, 278
Numerical distance, 74

Obstacle gain, 136
Optical frequencies, propagation in troposphere, 206, 336
Optimum traffic frequency, 311
Ordinary ray, 245, 247
absorption coefficient, 250
OTF, 311

Parabolic-equation method, 102
Path-length difference, optical, 147
Path loss, 24
PCA, 309
Permanent molecular dipole moment, 142
Permittivity, complex, 20
relative, 20
Phase shift in reflection, 31
Phase velocity, 238
Photo-detachment, 227
Planck's constant, 210
Plasma frequency, 244, 325
Polar cap absorption, 309
Polarization, direct wave, 48
gas molecules, 142
reflected waves, 48
Polarization ellipse, 66
Poynting's vector, 21
Power, minimum discernible, 38
Power flux density, 21
Precipitation statics, 337
Probable time of reliable reception, 177
Propagation, ground, 18
effects of earth, 21
in imperfectly conducting medium, 26-30
ionospheric, 18
line-of-sight, 93
trans-horizon, 173
tropospheric, 18
Pseudo-Brewster angle, 36

Radiation belts, 211
Radiation pattern, with ground reflections, 55
without ground reflections, 55
Radiation resistance, 72
dipole, 72
horizontal, 72
vertical, 72
Radiator, isotropic, 21
Radio circuit, 12
short-wave, calculation, 310
Radio communications, long-distance, 159
short-wave, conditions for, 293
Radio emission from sun, 348
Radio frequencies, spectrum of, 10
Radio link, 12
design principles, 37
tropospheric, approximate calculation, 191
Radio propagation, audio frequencies, 274
centimetric waves, 333
controlling factors, 14
decimetric waves, 333
effect of earth's curvature, 93-98, 118
effect of surface irregularities, 41
in homogeneous ionized gas, 235
in presence of magnetic field, 242
line-of-sight, 124
over rough surfaces, 129
long waves, 275
field-strength calculation, 281
signal-level variations, 279
medium waves, 284
signal-level variations, 285
metric waves, 323
ionospheric scatter, 323
meteor-trail, 328
MHD waves, 271
millimetric waves, 333
mixed-path, 81
modes of, table, 18
over flat earth, 70
over homogeneous earth, 84
over inhomogeneous spherical earth, 117
over inhomogeneous surface, 80
over knife edge, 133
over smooth homogeneous spherical earth, 100
over-the-horizon, 324
ray-tracing approach to, 157
significant volumes in, 118-19
short-wave, 290
effect of geomagnetic disturbances, 307
signal-level variations, 304
sub-audio frequencies, 271
trajectory of minimum absorption, 88
under conditions of super-refraction, 157
very long waves, 275
field-strength calculations, 281
signal-level variations, 279
Radio relay system, line-of-sight, 12
Radio spectrum, break-down of, 9
decimal classification of, 9
Soviet classification, 11
Radio waves, 9
attenuation in troposphere, 195
audio-frequency, 9
diffraction around spherical earth, 102
dispersion of, 39
lower limit, 9
reflection of, conditions for, 252
effect of earth magnetic field, 256
from earth, 31-6
in ionosphere, 250
refraction in ionosphere, 250
sub-audio frequency, 9
upper limit, 9
Rain, basic characteristics, 197
Ray, extraordinary, 245, 247
ordinary, 245, 247

Rayleigh ensemble, 184
Rayleigh-Jeans equation, 339
Rayleigh scattering, 200
Rayleigh's criterion, 129
Ray-tracing approach, 157
Reception, by ground aerial, 68
  diversity, 187
  frequency, 187
  space, 187
Recombination, 225
  effective, 227
  multi-step, 227
  simple, 227
Reflected ray, 58
Reflection, diffuse, 129
  specular, 129
Reflection coefficient, complex, 31
  effect of vegetation on, 131
  of rough surface, 130
Refraction, standard, 146
Refractive index, effect of turbulence, 161
Refractivity, total, 143
Refractometer, 154
Reliability of communication, 193
Reliable reception, probable time of, 177
Repeaters, radio-relay, 12
Rician density function, 184
Riometers, 344
Rotation of plane of polarization, 354

Satellite, active communication, 353, 357
  synchronous, 352, 357
  two-hour, 352
Satellite communication, general principles, 353
  systems, 356
    energy considerations, 356
    frequencies for, 356
Satellites, positioned, 359
  random-orbit, 359
Scale distance, 103
Scale height, 103, 213
Scatter, forward, 171
  ionospheric, 13, 17
    metric waves, 324
  meteor-trail, 13, 17
    metric waves, 328
  systems, 13
  tropospheric, 13, 159, 164
    statistical approach to, 170
Scattering, Mie's, 200
  Rayleigh, 200
Scattering anglo, 165
Scattering cross-section, 165
  determination of, 167
Scattering volume, 165
  determination, 173
  significant, 167
Semi-shadow zone, 101
Sferics, 337
Shadow zone, 101
Shchukin's method, 78
Shuleikin-van der Pol equation, 73, 75, 81
  limiting distances for, 77
SID, 281, 309
Signal level, seasonal variations, 188
Significant volumes in radio transmission, 118
Skip distance, 299
Skip zone, 300
  primary, 299
Snell's law, 250
Solar constant of radiation, 220
Solar flares, 219
Solar radiation, 218
Sommerfeld's solution, 73
Space communication, 351
Space wave, 17
Spectrum, radio frequencies, 10
Spherical-earth problem, 93
  practical solution of, 98
Spherical wave, 25
Sporadic *E* layer, 234, 270, 323
Standard pressure lapse rate, 141
Standard temperature lapse rate, 141

Standard troposphere, 141
Stellar radiation, 219
Sub-millimetre range, 9
Sub-refraction, 152
Sudden ionospheric disturbances, 281, 309
Sunspot cycle, 219, 281, 289, 304
Sunspot number, 304
Super-refraction, 152
  conditions for, 155
Surface-impedance conditions, 60

Tangent approximation, 102
Temperature inversions, 144, 155
  elevated, 156
Transmission, behind-the-horizon, 13
Transmission equation, ideal, 73
Transmission loss, 24
  basic, 24
Transmission, over-the-horizon, 13
Tropopause, 140
Troposphere, 16
  properties of, 140
  refractive index of, 16, 142
Tropospheric duct, 16, 158
Tropospheric propagation, at optical frequencies, 206
Tropospheric scatter link, uses for, 195
Tropospheric waveguide, 16, 158
Turbulence, 161
  local isotropy of, 192

Van Allen belts, 234
Van der Pol's equation, 73
Very long waves, uses for, 281
Virtual height, ionized layer, 259
Virtual ionospheric height, 263
Virtual ionospheric path, 263
Volume, common, 165
  scattering, 165
  significant, 167
Vvedensky's equation, 55

Watson's solution, 101
Wave, direct, 13, 14
    polarization of, 48
  extraordinary, 297
  ground, 15
  ionospheric, 17
  ordinary, 297
  polarized, 32
  reflected, 48
    polarization of, 48
  secondary, 122
    virtual source of, 122
  spherical, 25
  surface, 15
  tropospheric, 16
Wave field, spatial structure, 53
Wave front, tilt of, 68
Wave impedance, free-space, 21
Wave interference, 51
Wave number of air, 23
Wave path curvature, 145
Wave polarization, 31, 32, 48
  horizontal, 31, 48
  vertical, 31, 48
Wave propagation, guided, 159
  controlling factors, 14
  unguided, 11
    uses of, 11
Wave scattering, conditions for, 168
Waveguide, tropospheric, 16
Wavelengths, day-time, 293
  night-time, 293
Waves, magnetohydrodynamic, 9
Whisterlers, 274
Whistling atmospherics, 274
Windows for radio waves, 201, 203, 204, 333
Wolf number, 304
Work function, ionization, 213

Zenneck waves, 91
Zenneck's theory, 90
Zone of silence, 299

Лицензия ЛР № 063377 от 23.05.94г.
Формат 60X88/16. Печ. л. 23,5.
Тираж 1000. экз. Зак № 967.
Издательство "УРСС"
111672, г. Москва,
ул. Новокосинская 27/174

Московская тип. №4 Комитета РФ по печати
129041, Москва, Б. Переяславская, 46